Segmented compressed sampling for analog-to-information conversion: Method and performance analysis

Omid Taheri, *Student Member, IEEE,* and Sergiy A. Vorobyov, *Senior Member, IEEE*

Abstract

A new segmented compressed sampling method for analog-to-information conversion (AIC) is proposed. An analog signal measured by a number of parallel branches of mixers and integrators (BMIs), each characterized by a specific random sampling waveform, is first segmented in time into M segments. Then the sub-samples collected on different segments and different BMIs are reused so that a larger number of samples than the number of BMIs is collected. This technique is shown to be equivalent to extending the measurement matrix, which consists of the BMI sampling waveforms, by adding new rows without actually increasing the number of BMIs. We prove that the extended measurement matrix satisfies the restricted isometry property with overwhelming probability if the original measurement matrix of BMI sampling waveforms satisfies it. We also show that the signal recovery performance can be improved significantly if our segmented AIC is used for sampling instead of the conventional AIC. Simulation results verify the effectiveness of the proposed segmented compressed sampling method and the validity of our theoretical studies.

Index Terms

Compressed sampling, analog-to-information converter, correlated random variables, l_1-norm minimization, empirical risk minimization.

The authors are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada. The contacting emails are {otaheri, vorobyov}@ece.ualberta.ca.

Corresponding author: Sergiy A. Vorobyov, Dept. of Electrical and Computer Engineering, University of Alberta, 9107-116 St., Edmonton, Alberta, T6G 2V4, Canada; Phone: +1 (780) 492 9702, Fax: +1 (780) 492 1811.

This work was supported in parts by research grants from the Natural Science and Engineering Research Council (NSERC) of Canada and Alberta Ingenuity New Faculty Award.

I. INTRODUCTION

According to Shannon's sampling theorem, an analog band-limited signal can be recovered from its discrete-time samples if the sampling rate is at least twice the maximum frequency present in the signal. Recent theory of compressed sampling (CS), however, suggests that a signal can be recovered from fewer samples if it is sparse or compressible [1]–[4]. CS theory also suggests that a universal sampling matrix (for example, a random projection matrix) can be designed, and it can be used for all sparse signals regardless of their nature [2]. CS has already found a wide range of applications such as image acquisition [5], sensor networks [6], cognitive radios [7], communication channel estimation [8], [9], etc.

The sampling process often used in the CS literature consists of two steps. First, an analog signal is sampled at the Nyquist rate and then a measurement matrix is applied to the time domain samples in order to collect the compressed samples (see, for example, [7]). This sampling approach, however, defeats one of the primary purposes of CS, which is avoiding high rate sampling. A more practical approach for "direct" sampling and compression of analog signals has been presented in [10]. The analog signal is assumed to belong to the class of signals in shift-invariant spaces, that is, the analog signal can be represented as a linear combination of a set of m basis functions defined over a period T. The analog signal is first passed through a filter bank where each filter is matched to one of the m basis functions and the output is sampled at time instances nT where n is an integer. If the signal is sparse, then only $S < m$ samples are nonzero. The set of m output samples are then passed through a measurement matrix to create $K \geq S$ compressed samples representing the analog signal in a specific period $[(n-1)T,\, nT]$. It is worth mentioning that this method is a generalization of another method in [11] which is devised for sub-Nyquist sampling of multi-band signals. The limits of this method come from the underlying assumption that the signal belongs to the class of signals in shift-invariant spaces. Although this assumption is argued to be valid for a variety of engineering applications [10], [12] and can be generalized to the signals in a union of subspaces [13], [14], it is still a limiting assumption. Moreover, the complexity of this method is by no means lower than the complexity of another practical approach to CS, which avoids high rate sampling [1], [15]. The name analog-to-information converter (AIC) has been coined for the latter method. The AIC consists of several parallel branches of mixers and integrators (BMIs) in which the analog signal is measured against different random sampling waveforms. Therefore, for every collected compressed sample, there is a BMI that multiplies the signal to a sampling waveform and then integrates the result over a period T.

In this paper, we propose a new segmented AIC structure with the goal of reducing the hardware

complexity.[1] The contributions of this work are the following. (i) A new segmented AIC structure is developed. In this structure, the integration period T is divided into M equal subperiods such that the sampling rate of our segmented AIC scheme is M times higher than of the AIC of [1]. The sub-samples collected over different subperiods by combining the sub-samples from different BMIs are then reused in order to build additional samples. In this way, a number of samples larger than the number of BMIs can be collected, although such samples will be correlated. We show that our segmented AIC technique is equivalent to extending the measurement matrix which consists of the BMI sampling waveforms by adding new rows without actually increasing the number of BMIs. In this respect, the following works also need to be mentioned [17], [18]. In [17], Toeplitz-structured measurement matrices are considered, while measurement matrices built on one random vector with shifts of $D \geq 1$ in between the rows appear in radar imaging application considered in [18]. (ii) We show that the restricted isometry property (RIP), that is a sufficient condition for signal recovery based on compressed samples, is satisfied for the extended measurement matrix resulting from the segmented AIC structure with overwhelming probability if the original matrix of BMI sampling waveforms satisfies the RIP. Thus, our segmented AIC is a valid candidate for CS. (iii) We also show that the signal recovery performance improves if our segmented AIC is used for sampling instead of the AIC of [1] with the same number of BMIs. The mathematical challenge in this part of the work is that the samples collected by our segmented AIC are correlated, while all available results on performance analysis of the signal recovery are obtained for the case of uncorrelated samples.

The rest of this paper is organized as follows. Necessary background on CS, CS signal recovery, and AIC is briefly summarized in Section II. The main idea of the paper, that is, the segmented AIC structure, is explained in Section III. We prove in Section IV that the extended measurement matrix resulting from the proposed segmented AIC satisfies the RIP and, therefore, the segmented AIC is a legitimate CS method. The signal recovery performance analysis for our segmented AIC is summarized in Section V. Section VI demonstrates the simulation results and Section VII concludes the paper.

II. Background

CS basics and notations: CS deals with a low rate representation of sparse signals, i.e., such signals which have few nonzero projections on the vectors of an orthogonal basis (sparsity basis). Let $\mathbf{\Psi} = \left(\boldsymbol{\psi}_1^T, \boldsymbol{\psi}_2^T, \ldots, \boldsymbol{\psi}_N^T\right)^T$ be an $N \times N$ matrix of basis vectors $\boldsymbol{\psi}_i,\ i = 1, \ldots, N$, i.e., the sparsity basis, and

[1]Some preliminary results have been reported in [16].

$\boldsymbol{f}$ be a discrete-time sparse signal[2] represented in this basis as

$$\boldsymbol{f} = \sum_{i=1}^{N} x_i \boldsymbol{\psi}_i^H = \boldsymbol{\Psi}^H \boldsymbol{x} \tag{1}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)^T$ is the $N \times 1$ vector of coefficients and $(\cdot)^T$ and $(\cdot)^H$ stand for the transpose and Hermitian transpose, respectively. A signal is S-sparse if at most S projections on the rows of $\boldsymbol{\Psi}$, i.e., coefficients of $\boldsymbol{x}$, are nonzero. It is known that a universal compressed sampling method can be designed to effectively sample and recover S-sparse signals regardless of the specific sparsity domain [1], [2].

Among various bounds on the sufficient number of collected compressed samples[3] K ($S < K < N$) required for recovering an S-sparse signal, the first and most popular one is given by the following inequality $S \leq CK/\log(N/K)$ where C is some constant [1]. This bound is derived based on the uniform uncertainty principle [20]. Let $\boldsymbol{\Phi}$ be a $K \times N$ measurement matrix applied to a sparse signal for collecting K compressed samples. Then the uniform uncertainty principle states that $\boldsymbol{\Phi}$ must satisfy the following restricted isometry property (RIP) [1]. Let $\boldsymbol{\Phi}_{\mathcal{T}}$ be a sub-matrix of $\boldsymbol{\Phi}$ retaining only the columns with their indexes in the set $\mathcal{T} \subset \{1, \ldots, N\}$. Then the S-restricted isometry constant δ_S is the smallest number satisfying the inequality

$$\frac{K}{N}(1 - \delta_S)\|\boldsymbol{c}\|_{l_2}^2 \leq \|\boldsymbol{\Phi}_{\mathcal{T}}\boldsymbol{c}\|_{l_2}^2 \leq \frac{K}{N}(1 + \delta_S)\|\boldsymbol{c}\|_{l_2}^2 \tag{2}$$

for all sets $\mathcal{T}$ of cardinality less than or equal to S and all vectors $\boldsymbol{c}$ (here $\|\cdot\|_{l_2}$ denotes the Euclidean norm of a vector). As shown in [2], [21], if the entries of $\boldsymbol{\Phi}$ are, for example, independent zero mean Gaussian variables with variance $1/N$, then $\boldsymbol{\Phi}$ satisfies the RIP for $S \leq CK/\log(N/K)$ with high probability.[4]

Recovery methods: Using the measurement matrix $\boldsymbol{\Phi}$, the $1 \times K$ vector of compressed samples $\boldsymbol{y}$ can be calculated as $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{f} = \boldsymbol{\Phi}'\boldsymbol{x}$ where $\boldsymbol{\Phi}' = \boldsymbol{\Phi}\boldsymbol{\Psi}^H$. A signal can be recovered from its noiseless sample vector $\boldsymbol{y}$ based on the following convex optimization problem that can be solved by a linear program [2], [22]

$$\min\|\tilde{\boldsymbol{x}}\|_{l_1} \quad \text{subject to} \quad \boldsymbol{\Phi}'\tilde{\boldsymbol{x}} = \boldsymbol{y} \tag{3}$$

where $\|\cdot\|_{l_1}$ denotes the l_1-norm of a vector.

[2]It can be in $\mathbb{R}^N$ or $\mathbb{C}^N$.

[3]See [19] for broader review.

[4]Note that in order to ensure consistency throughout the paper, the variance of the elements in $\boldsymbol{\Phi}$ is taken to be $1/N$ instead of $1/K$ as, for example, in [2]. Thus, the multiplier K/N is added in the left- and right-hand sides of (2).

If the compressed samples are noisy, the sampling process can be expressed as

$$\boldsymbol{y} = \boldsymbol{\Phi f} + \boldsymbol{w} \tag{4}$$

where $\boldsymbol{w}$ is a zero mean noise vector with identically and independently distributed (i.i.d.) entries of variance σ^2. Then the recovery problem is modified as [23]

$$\min\|\tilde{\boldsymbol{x}}\|_{l_1} \quad \text{subject to} \quad \|\boldsymbol{\Phi}^{'}\tilde{\boldsymbol{x}} - \boldsymbol{y}\|_{l_2} \leq \gamma \tag{5}$$

where γ is the bound on the square root of the noise energy.

Another technique for sparse signal recovery from noisy samples (see [4]) uses the empirical risk minimization method that was first developed in statistical learning theory for approximating an unknown function based on noisy measurements [24]. Note that the empirical risk minimization-based recovery method is of a particular interest since under some simplifications (see [4, p. 4041]) it reduces to another well-known least absolute shrinkage and selection operator (LASSO) method [25]. Therefore, the risk minimization-based method of [4] provides the generality which we need in this paper.

In application to CS, the unknown function is the sparse signal and the noisy compressed samples are the collected data. Let the entries of the measurement matrix $\boldsymbol{\Phi}$ be selected with equal probability as $\pm 1/\sqrt{N}$, and the energy of the signal $\boldsymbol{f}$ be bounded so that $\|\boldsymbol{f}\|^2 \leq NB^2$. The risk $r(\hat{\boldsymbol{f}})$ of a candidate reconstruction $\hat{\boldsymbol{f}}$ and its empirical risk $\hat{r}(\hat{\boldsymbol{f}})$ are defined as follows [24]

$$r(\hat{\boldsymbol{f}}) = \frac{\|\hat{\boldsymbol{f}} - \boldsymbol{f}\|^2}{N} + \sigma^2, \qquad \hat{r}(\hat{\boldsymbol{f}}) = \frac{1}{K}\sum_{j=1}^{K}\left(y_j - \boldsymbol{\phi}_j\hat{\boldsymbol{f}}\right)^2. \tag{6}$$

Then the candidate reconstruction $\hat{\boldsymbol{f}}_K$ obtained based on K samples can be found as [4]

$$\hat{\boldsymbol{f}}_K = \arg\min_{\hat{\boldsymbol{f}}\in\mathcal{F}(B)}\left\{\hat{r}(\hat{\boldsymbol{f}}) + \frac{c(\hat{\boldsymbol{f}})\log 2}{\epsilon K}\right\} \tag{7}$$

where $\mathcal{F}(B) = \{\boldsymbol{f} : \|\boldsymbol{f}\|^2 \leq NB^2\}$, $c(\hat{\boldsymbol{f}})$ is a nonnegative number assigned to a candidate signal $\hat{\boldsymbol{f}}$, and $\epsilon = 1/\left(50(B+\sigma)^2\right)$. Moreover, $\hat{\boldsymbol{f}}_K$ given by (7) satisfies the following inequality [4]

$$E\left\{\frac{\|\hat{\boldsymbol{f}}_K - \boldsymbol{f}\|^2}{N}\right\} \leq C_1 \min_{\hat{\boldsymbol{f}}\in\mathcal{F}(B)}\left\{\frac{\|\hat{\boldsymbol{f}} - \boldsymbol{f}\|^2}{N} + \frac{c(\hat{\boldsymbol{f}})\log 2 + 4}{\epsilon K}\right\} \tag{8}$$

where $C_1 = [(27-4e)(B/\sigma)^2 + (50-4\sqrt{2})B/\sigma + 26]/[(23-4e)(B/\sigma)^2 + (50-4\sqrt{2})B/\sigma + 24]$, $e = 2.7183\ldots$, and $E\{\cdot\}$ stands for the expectation operation.

Let a compressible signal $\boldsymbol{f}$ be defined as a signal for which $\|\boldsymbol{f}^{(m)} - \boldsymbol{f}\|^2 \leq NC_A m^{-2\alpha}$ where $\boldsymbol{f}^{(m)}$ is the best m-term approximation of $\boldsymbol{f}$ which is obtained by retaining the m most significant coefficients of vector $\boldsymbol{x}$ ($\boldsymbol{x}$ being the representation of $\boldsymbol{f}$ in the sparsity basis $\boldsymbol{\Psi}$), and $C_A > 0$ and $\alpha \geq 0$ are

some constants. Let also $\mathcal{F}_c(B,\alpha,C_A)=\{\boldsymbol{f}:\|\boldsymbol{f}\|^2\le NB^2,\|\boldsymbol{f}^{(m)}-\boldsymbol{f}\|^2\le NC_Am^{-2\alpha}\}$ be the set of compressible signals. Then based on the weight assignment $c(\boldsymbol{f})=2\log(N)N_{\boldsymbol{x}}$ (here $N_{\boldsymbol{x}}$ is the actual number of nonzero coefficients in $\boldsymbol{x}$) the following inequality holds [4]

$$\sup_{\boldsymbol{f}\in\mathcal{F}_c(B,\alpha,C_A)} E\left\{\frac{\|\hat{\boldsymbol{f}}_K-\boldsymbol{f}\|^2}{N}\right\}\le C_1C_2\left(\frac{K}{\log N}\right)^{-2\alpha/(2\alpha+1)} \tag{9}$$

where $C_2=C_2(B,\sigma,C_A)>0$ is a constant.

If signal $\boldsymbol{f}$ is indeed sparse and belongs to $\mathcal{F}_s(B,S)=\{\boldsymbol{f}:\|\boldsymbol{f}\|^2\le NB^2,\|\boldsymbol{f}\|_{l_0}\le S\}$, then there exists a constant $C_2'=C_2'(B,\sigma)>0$ such that [4]

$$\sup_{\boldsymbol{f}\in\mathcal{F}_s(B,S)} E\left\{\frac{\|\hat{\boldsymbol{f}}_K-\boldsymbol{f}\|^2}{N}\right\}\le C_1C_2'\left(\frac{K}{S\log N}\right)^{-1}. \tag{10}$$

AIC: The random modulation preintegration (RMPI) structure is proposed for AIC in [1]. The RMPI multiplies the signal and the sampling waveforms in the analog domain and then integrates the product over the signal period to produce samples. It implies that the sampling device has a number of parallel BMIs in order to process the analog signal in real-time. The RMPI structure is shown in Fig. 1, where $f(t)$ is the analog signal being sampled, $\phi_i(t),\ i=1,\dots,K$ are the sampling waveforms (rows of the measurement matrix $\mathbf{\Phi}$), and $y_i,\ i=1,\dots,K$ are the compressed samples.

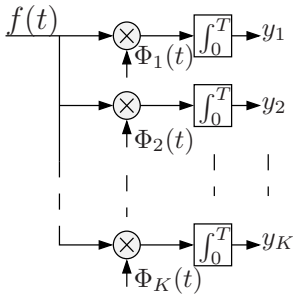


Fig. 1. The structure of the AIC based on RMPI.

III. Segmented Compressed Sampling Method

AIC removes the need for high speed sampling, but it may still be necessary in many practical applications to collect a larger number of compressed samples than the AIC hardware (the number of parallel BMIs) may allow. Indeed, a smaller number of samples may have a negative effect on the signal recovery accuracy which can be an issue in a number of applications. In order to collect a larger number of compressed samples using AIC, we need to increase the hardware complexity by adding more

BMIs. The latter makes the AIC device complex and expensive although its sampling rate is much lower than that of analog-to-digital converter (ADC). Therefore, it is desirable to reduce the number of parallel BMIs in AIC without sacrificing the signal recovery accuracy. It can be achieved by adding to AIC the capability of sampling at a higher rate, which is, however, significantly lower than the sampling rate required by ADC. The latter can be achieved by splitting the integration period T in every BMI of the AIC in Fig. 1 into shorter subperiods. It is equivalent to generating a number of incomplete samples of a signal. Note that since the original integration period is divided into a number of smaller subperiods, the samples collected over all parallel BMIs during one subperiod do not have complete information about the signal. Therefore, they are called incomplete samples. Hereafter, the complete samples obtained over the whole period T are referred to as just samples, while the incomplete samples are referred to as sub-samples.

A. The Basic Idea and the Model

The basic idea is to collect the sub-samples as described above and then reuse them in order to build additional samples. In this manner, a larger number of samples than the number of BMIs can be collected. It allows for a tradeoff between AIC and ADC since as in AIC the signal is measured at a low rate by correlating it to a number of sampling waveforms, while the integration period is split into shorter sub-intervals which is similar to the requirement of a higher sampling rate as in ADC. However, the required sampling rate in the proposed scheme is still significantly lower than that required by ADC.

Let the integration period be split into M sub-intervals, and let $\boldsymbol{y}_k = (y_{k,1},\ \ldots,\ y_{k,M})^T,\ k = 1, \ldots, K$ be the vectors of sub-samples collected against the sampling waveforms $\boldsymbol{\phi}_k,\ k = 1, \ldots, K$, where K is the original number of sampling waveforms, i.e., the number of BMIs. The sub-sample $y_{k,j}$ is given by

$$y_{k,j} = \int_{(j-1)T/M}^{jT/M} x(t)\boldsymbol{\phi}_k(t)dt. \tag{11}$$

Then the total number of sub-samples collected in all BMIs over all subperiods is MK. These sub-samples can be gathered in the following $K \times M$ matrix

$$\boldsymbol{Y} = \begin{pmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,M} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,M} \\ \vdots & \vdots & \vdots & \vdots \\ y_{K,1} & y_{K,2} & \cdots & y_{K,M} \end{pmatrix} \tag{12}$$

where the k-th row contains the sub-samples obtained by correlating the measured signal with the waveform $\boldsymbol{\phi}_k$ over M subperiods each of length T/M.

The original K samples, i.e., the samples collected at BMIs over the whole time period T, are

$$y_k = \sum_{m=1}^{M} [\boldsymbol{Y}]_{k,m}, \qquad k = 1, \ldots, K \tag{13}$$

where $[\boldsymbol{Y}]_{k,m}$ denotes the (k,m)-th element of $\boldsymbol{Y}$, that is, $[\boldsymbol{Y}]_{k,m} = y_{k,m}$.

In order to construct additional samples to the samples obtained using (13), we consider columnwise permuted versions of $\boldsymbol{Y}$. The following definitions are then in order.

The *permutation* π is a one-to-one mapping of the elements of a set $\mathcal{D}$ to itself by simply changing the order of the elements. Then $\pi(k)$ stands for the index of the k-th element in the permuted set. For example, let $\mathcal{D}$ consists of the elements of a $K \times 1$ vector $\boldsymbol{z}$, and the order of the elements in $\mathcal{D}$ is the same as in $\boldsymbol{z}$. After applying the permutation function π to $\boldsymbol{z}$, the permuted vector is $\boldsymbol{z}^{\pi} = \left(z_{\pi(1)}, \ldots, z_{\pi(k)}, \ldots, z_{\pi(K)}\right)^T$. If vector $\boldsymbol{z}$ is itself the vector of indexes, i.e., $\boldsymbol{z} = (1, \ldots, K)^T$, then obviously $z_{\pi(k)} = \pi(k)$.

The permuted versions of the sub-sample matrix $\boldsymbol{Y}$ can be obtained by applying different permutations to different columns of $\boldsymbol{Y}$. Specifically, let $\mathcal{P}^{(i)} = \{\pi_1^{(i)}, \ldots, \pi_j^{(i)}, \ldots, \pi_M^{(i)}\}$ be the i-th set of column permutations with $\pi_j^{(i)}$ being the permutation function applied to the j-th column of $\boldsymbol{Y}$, and let I stand for the number of such permutation sets. Then according to the above notations, the matrix resulting from applying the set of permutations $\mathcal{P}^{(i)}$ to the columns of $\boldsymbol{Y}$ can be expressed as $\boldsymbol{Y}^{\mathcal{P}^{(i)}} = \left(\boldsymbol{y}_1^{\pi_1^{(i)}}, \ldots, \boldsymbol{y}_j^{\pi_j^{(i)}}, \ldots, \boldsymbol{y}_M^{\pi_M^{(i)}}\right)$ where $\boldsymbol{y}_j$ is the j-th column of $\boldsymbol{Y}$.

Permutation sets $\mathcal{P}^{(i)},\ i = 1, \ldots, I$ are chosen in such a way that all sub-samples in a specific row of $\boldsymbol{Y}^{\mathcal{P}^{(i)}}$ come from different rows of the original sub-sample matrix $\boldsymbol{Y}$ as well as from different rows of other permuted matrices $\boldsymbol{Y}^{\mathcal{P}^{(1)}}, \ldots, \boldsymbol{Y}^{\mathcal{P}^{(i-1)}}$. For example, all sub-samples in a specific row of $\boldsymbol{Y}^{\mathcal{P}^{(1)}}$ must come from different rows of the original matrix $\boldsymbol{Y}$ only, while the sub-samples in a specific row of $\boldsymbol{Y}^{\mathcal{P}^{(2)}}$ come from different rows of $\boldsymbol{Y}$ and $\boldsymbol{Y}^{\mathcal{P}^{(1)}}$ and so on. This requirement is forced to make sure that any additional sample has the least possible correlation with the original samples of (13). Then the additional $K\,I$ samples can be obtained based on the permuted matrices $\boldsymbol{Y}^{\mathcal{P}^{(i)}},\ i = 1, \ldots, I$ as

$$y_k^{\mathcal{P}^{(i)}} = \sum_{m=1}^{M} [\boldsymbol{Y}^{\mathcal{P}^{(i)}}]_{k,m}, \qquad k = 1, \ldots, K \quad i = 1, \ldots, I. \tag{14}$$

It is worth noting that in terms of the hardware structure, the sub-samples used to generate additional samples must be chosen from different BMIs as well as different integration subperiods. This is equivalent to collecting additional samples by correlating the signal with additional sampling waveforms which are not present among the actual BMI sampling waveforms. Each of these additional sampling waveforms comprises the non-overlapping subperiods of M different original waveforms.

Now the question is how many permuted matrices, which satisfy the above summarized conditions, can be generated based on $\boldsymbol{Y}$. Consider the following $K \times M$ matrix

$$\boldsymbol{Z} \triangleq \underbrace{(\boldsymbol{z}, \boldsymbol{z}, \ldots, \boldsymbol{z})}_{M \text{ times}} \tag{15}$$

where $\boldsymbol{z}$ is the vector of indexes. Applying the column permutation set $\mathcal{P}^{(i)}$ to the columns of $\boldsymbol{Z}$, we obtain a permuted matrix $\boldsymbol{Z}^{\mathcal{P}^{(i)}} = \left(\boldsymbol{z}^{\pi_1^{(i)}}, \ldots, \boldsymbol{z}^{\pi_j^{(i)}}, \ldots, \boldsymbol{z}^{\pi_M^{(i)}}\right)$. Then the set of all permuted versions of $\boldsymbol{Z}$ can be denoted as $\mathcal{S}_{\boldsymbol{Z}} = \{\boldsymbol{Z}^{\mathcal{P}^{(1)}}, \ldots, \boldsymbol{Z}^{\mathcal{P}^{(I)}}\}$. With these notations, the following theorem is in order.

Theorem 1. *The size of $\mathcal{S}_{\boldsymbol{Z}}$, i.e., the number I of permutation sets $\mathcal{P}^{(i)},\ i = 1, \ldots, I$ which satisfy the conditions*

$$[\boldsymbol{Z}^{\mathcal{P}^{(i)}}]_{k,j} \neq [\boldsymbol{Z}^{\mathcal{P}^{(i)}}]_{k,r},\ \forall \boldsymbol{Z}^{\mathcal{P}^{(i)}} \in \mathcal{S}_{\boldsymbol{Z}},\ \ j \neq r,\ k \in \{1, \ldots, K\},\ j, r \in \{1, \ldots, M\} \tag{16}$$

$$\exists! j \text{ or } \nexists j \text{ such that } [\boldsymbol{Z}^{\mathcal{P}^{(i)}}]_{k,j} = [\boldsymbol{Z}^{\mathcal{P}^{(l)}}]_{h,j}, \quad \forall \boldsymbol{Z}^{\mathcal{P}^{(i)}}, \boldsymbol{Z}^{\mathcal{P}^{(l)}} \in \mathcal{S}_{\boldsymbol{Z}},\ \boldsymbol{Z}^{\mathcal{P}^{(i)}} \neq \boldsymbol{Z}^{\mathcal{P}^{(l)}}, \forall j \in \{1, \ldots, M\}$$
$$\forall k, h \in \{1, \ldots, K\} \tag{17}$$

is at most $K - 1$. Here $[\boldsymbol{Z}^{\mathcal{P}^{(i)}}]_{k,j}$ stands for the (k, j)-th element of the permuted matrix $\boldsymbol{Z}^{\mathcal{P}^{(i)}}$.

Remark 1. *Using the property that $z_{\pi(k)} = \pi(k)$ for the vector of indexes $\boldsymbol{z}$, the conditions* (16) *and* (17) *can also be expressed in terms of permutations as*

$$\pi_j^{(i)}(k) \neq \pi_r^{(i)}(k) \quad \forall i \in \{1, \ldots, I\},\ j \neq r,\ k \in \{1, \ldots, K\},\ j, r \in \{1, \ldots, M\} \tag{18}$$

$$\exists! j \text{ or } \nexists j \text{ such that } \pi_j^{(i)}(k) = \pi_j^{(l)}(h)\ \forall i, l \in \{1, \ldots, I\},\ i \neq l,\ \forall j \in \{1, \ldots, M\}, \forall k, h \in \{1, \ldots, K\}. \tag{19}$$

Proof: See Appendix A. ■

Example 1: Let the specific choice of index permutations be $\pi_s(k) = ((s + k - 2) \bmod K) + 1,\ s, k = 1, \ldots, K$ with π_1 being the identity permutation and 'mod' standing for the modulo operation. For this specific choice, $\pi_j^{(i)} = \pi_{[i(j-1) \bmod K]+1},\ i = 1, \ldots, K-1,\ j = 1, \ldots, M$. Consider the following matrix notation for the set $\mathcal{P}$ where the elements along the i-th row are the permutations $\mathcal{P}^{(i)}$, $i = 1, \ldots, I$

$$\mathcal{P} \triangleq \begin{pmatrix} \mathcal{P}^{(1)} \\ \mathcal{P}^{(2)} \\ \mathcal{P}^{(3)} \\ \vdots \\ \mathcal{P}^{(K-2)} \\ \mathcal{P}^{(K-1)} \end{pmatrix} = \begin{pmatrix} \pi_1^{(1)} & \pi_2^{(1)} & \pi_3^{(1)} & \dots & \pi_M^{(1)} \\ \pi_1^{(2)} & \pi_2^{(2)} & \pi_3^{(2)} & \dots & \pi_M^{(2)} \\ \pi_1^{(3)} & \pi_2^{(3)} & \pi_3^{(3)} & \dots & \pi_M^{(3)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \pi_1^{(K-2)} & \pi_2^{(K-2)} & \pi_3^{(K-2)} & \dots & \pi_M^{(K-2)} \\ \pi_1^{(K-1)} & \pi_2^{(K-1)} & \pi_3^{(K-1)} & \dots & \pi_M^{(K-1)} \end{pmatrix}$$

$$= \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \dots & \pi_M \\ \pi_1 & \pi_3 & \pi_5 & \dots & \pi_{[2(M-1) \bmod K]+1} \\ \pi_1 & \pi_4 & \pi_7 & \dots & \pi_{[3(M-1) \bmod K]+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_{K-1} & \pi_{K-3} & \dots & \pi_{[(K-2)(M-1) \bmod K]+1} \\ \pi_1 & \pi_K & \pi_{K-1} & \dots & \pi_{[(K-1)(M-1) \bmod K]+1} \end{pmatrix}. \tag{20}$$

Note that not all permutations $\mathcal{P}^{(i)}$, $i = 1, \dots, I$ used in (20) may be permissible. In fact, the set of permutations $\mathcal{P}^{(i)}$ with $K/gcd(i, K) < M$ has at least one repeated permutation that contradicts the condition (18). Here $gcd(\cdot, \cdot)$ stands for the greatest common devisor of two numbers. For example, for $K = 8$ and $M = 4$, $K/gcd(4, K) = 2 < M$ and $\mathcal{P}^{(4)}$ is impermissible. Therefore, instead of $K - 1 = 7$, only the following 6 sets of permutations are allowed

$$\mathcal{P} = \begin{pmatrix} \pi_1^{(1)} & \pi_2^{(1)} & \pi_3^{(1)} & \pi_4^{(1)} \\ \pi_1^{(2)} & \pi_2^{(2)} & \pi_3^{(2)} & \pi_4^{(2)} \\ \pi_1^{(3)} & \pi_2^{(3)} & \pi_3^{(3)} & \pi_4^{(3)} \\ \pi_1^{(4)} & \pi_2^{(4)} & \pi_3^{(4)} & \pi_4^{(4)} \\ \pi_1^{(5)} & \pi_2^{(5)} & \pi_3^{(5)} & \pi_4^{(5)} \\ \pi_1^{(6)} & \pi_2^{(6)} & \pi_3^{(6)} & \pi_4^{(6)} \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \\ \pi_1 & \pi_3 & \pi_5 & \pi_7 \\ \pi_1 & \pi_4 & \pi_7 & \pi_2 \\ \pi_1 & \pi_6 & \pi_3 & \pi_8 \\ \pi_1 & \pi_7 & \pi_5 & \pi_3 \\ \pi_1 & \pi_8 & \pi_7 & \pi_6 \end{pmatrix}. \tag{21}$$

Theorem 1 shows how many different permuted versions of the original sub-sample matrix $\boldsymbol{Y}$ can be obtained such that the correlation between the original and additional samples would be minimal. Indeed, since the set of sub-samples that are used to build additional samples is chosen in a way that additional samples have at most one sub-sample in common with the previous samples, i.e., conditions (18) and (19) are satisfied, the set of permutations (20) is a valid candidate. The i-th element of $\mathcal{P}$, i.e., the element $\mathcal{P}^{(i)} = \left(\pi_1^{(i)}, \dots, \pi_M^{(i)}\right)$, is the set of permutations applied to $\boldsymbol{Y}$ to obtain $\boldsymbol{Y}^{\mathcal{P}^{(i)}}$. Adding up the entries along the rows of $\boldsymbol{Y}^{\mathcal{P}^{(i)}}$, a set of K additional samples can be obtained.

Example 2: Let the number of new samples K_a be at most K. This means that all permutations are given by only $\mathcal{P}^{(1)}$ in (20). In this special case, the sub-sample selection method can be summarized as follows. For constructing the $(K+1)$-st sample, M sub-samples on the main diagonal of $\boldsymbol{Y}$ are summed up together. Then the M sub-samples on the second diagonal are used to construct the $(K+2)$-nd sample, and so on up to the K_a-th sample. Mathematically, the so constructed additional samples can be expressed in terms of the elements of $\boldsymbol{Y}$ as

$$y_{K+k} = \sum_{m=1}^{M} y_{l,m}, \qquad k = 1 \ldots, K_a \tag{22}$$

where $l = [(k+m-2) \bmod K] + 1$ and $K_a \leq K$. Fig. 2 shows schematically how the sub-samples are selected in this example.

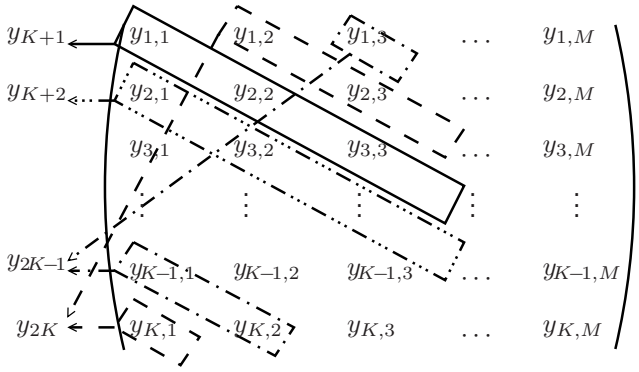


Fig. 2. Sub-sample selection principle for building additional samples in Example 2.

Our segmented sampling process can be equivalently expressed in terms of the measurement matrix. Let $\boldsymbol{\Phi}$ be the original $K \times N$ measurement matrix. Let the k-th row of the matrix $\boldsymbol{\Phi}$ be $\boldsymbol{\phi}_k = \left(\boldsymbol{\phi}_{k,1}, \ldots, \boldsymbol{\phi}_{k,M}\right)$ where $\boldsymbol{\phi}_{k,j},\ j = 1, \ldots, M$ are some vectors. Let for simplicity, the length of $\boldsymbol{\phi}_{k,j}$ be N/M and N/M be an integer number. The set of permutations applied to $\boldsymbol{Y}$ in order to obtain $\boldsymbol{Y}^{\mathcal{P}^{(i)}}$ is $\mathcal{P}^{(i)}$. Then the operation $\boldsymbol{\Phi}^{\mathcal{P}^{(i)}}$ can be expressed as follows. The first N/M columns of $\boldsymbol{\Phi}$, which are the vectors $\boldsymbol{\phi}_{k,1},\ k \in \{1, ..., K\}$, are permuted with $\pi_1^{(i)}$. The second N/M columns of $\boldsymbol{\Phi}$ are permuted with $\pi_2^{(i)}$ and so on until the last N/M columns of $\boldsymbol{\Phi}$ which are permuted with $\pi_M^{(i)}$. Then the extended measurement matrix which combines all possible permutations $\mathcal{P}^{(i)},\ i = 1, \ldots, I$ can be expressed as

$$\boldsymbol{\Phi}_e = \left(\boldsymbol{\Phi}^T, (\boldsymbol{\Phi}^{\mathcal{P}^{(1)}})^T, \ldots, (\boldsymbol{\Phi}^{\mathcal{P}^{(I)}})^T\right)^T \tag{23}$$

where $K_e \triangleq K + K_a = K + KI$.

Example 3: Continuing with the set up used in Example 2, let $K_a \leq K$. Then the extended measurement matrix is

$$\mathbf{\Phi}_e = \begin{pmatrix} \mathbf{\Phi} \\ \mathbf{\Phi}_1 \end{pmatrix} = \begin{pmatrix} \phi_{1,1} & \phi_{1,2} & \cdots & \phi_{1,M} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{K,1} & \phi_{K,2} & \cdots & \phi_{K,M} \\ \phi_{1,1} & \phi_{2,2} & \cdots & \phi_{M,M} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{K_a,1} & \phi_{\pi_2(K_a),M} & \cdots & \phi_{\pi_M(K_a),M} \end{pmatrix} \tag{24}$$

where $\mathbf{\Phi}_1$ contains only K_a rows of $\mathbf{\Phi}^{\mathcal{P}^{(1)}}$ and $\mathbf{\Phi}_1 = \mathbf{\Phi}^{\mathcal{P}^{(1)}}$ if $K_a = K$.

B. Implementation Issues and Discussion

Due to the special structure of the extended measurement matrix $\mathbf{\Phi}_e$, the sampling hardware needs only K parallel BMIs for collecting KI samples. These BMIs are essentially the same as those in Fig. 1. The only difference is that the integration period T is divided into M equal subperiods. After every subperiod, each integrator's output is sampled and the integrator is reset. In addition, a multiplexer which selects the sub-samples for constructing additional samples is needed. Note that partial sums can be kept for constructing the samples (original and additional), that is, the results of the integration are updated and accumulated for each sample iteratively after each subperiod. In this way, there is no need of designing the circuitry to memorize the matrix of sub-samples $\boldsymbol{Y}$, but only the partial sums for each sample are memorized at any current subperiod.

Since the proposed segmented AIC scheme collects the sub-samples at the M times higher rate than the AIC in Fig. 1, an improved signal recovery performance is expected. It agrees with the convention that the recovery performance cannot be improved only due to the post processing. Moreover, note that since the original random sampling waveforms are linearly independent with high probability, the additional sampling waveforms of our segmented compressed sampling method are also linearly independent with overwhelming probability. However, a sufficient condition that guarantees that the extended measurement matrix of the proposed segmented AIC scheme is an eligible choice is the RIP. Therefore, the RIP for the proposed segmented compressed sampling scheme is analyzed in the next section.

IV. RIP FOR THE SEGMENTED COMPRESSED SAMPLING METHOD

The purpose of this section is to show that the extended measurement matrix $\mathbf{\Phi}_e$ in (23) satisfies the RIP if the original measurement matrix $\mathbf{\Phi}$ satisfies it. The latter will also imply that $\mathbf{\Phi}_e$ can be used

as a valid CS measurement matrix. In our set up it is only assumed that the elements of the original measurement matrix are i.i.d. zero mean Gaussian variables and the measurement matrix is extended by adding its permuted versions as described in the previous section.

Let us first consider the special case of Example 3. In this case, $\mathbf{\Phi}$, $\mathbf{\Phi}_1$, and $\mathbf{\Phi}_e$ are the original measurement matrix, the matrix of additional sampling waveforms, and the extended measurement matrix given by (24), respectively. Let the matrix $\mathbf{\Phi}$ satisfy the RIP with sufficiently high probability. For example, let the elements of $\mathbf{\Phi}$ be i.i.d. zero mean Gaussian random variables with variance $1/N$. Let $\mathcal{T}$ be any subset of size S of the set $\{1,\ldots,N\}$. Then for any $0<\delta_S<1$, the matrix $\mathbf{\Phi}_{\mathcal{T}}$, which is a sub-matrix of $\mathbf{\Phi}$ which consists of only the columns with their indexes in the set $\mathcal{T}$ satisfies (2) with the following probability [21]

$$\Pr\{\mathbf{\Phi}_{\mathcal{T}} \text{ satisfies (2)}\} \geq 1-2\left(12/\delta_S\right)^S e^{-C_0(\delta_S/2)K} \tag{25}$$

where $C_0\left(\delta_S/2\right)=\delta_S^2/16-\delta_S^3/48$. Hereafter, the notation C_0 is used instead of $C_0\left(\delta_S/2\right)$ for brevity.

First, the following auxiliary result on the extended measurement matrix $\mathbf{\Phi}_e$ is of interest.

Lemma 1. *Let the elements of the measurement matrix $\mathbf{\Phi}$ be i.i.d. zero mean Gaussian variables with variance $1/N$, $\mathbf{\Phi}_e$ be formed as shown in* (24)*, and $\mathcal{T}\subset\{1,\ldots,N\}$ of size S. If K_a is chosen such that min$\{K, K_a+M-1\}\leq\lceil(K+K_a)/2\rceil$, then for any $0<\delta_S<1$, the following inequality holds*

$$Pr\{(\mathbf{\Phi}_e)_{\mathcal{T}} \text{ satisfies (2)}\} \geq 1-4\left(12/\delta_S\right)^S e^{-C_0\lfloor\frac{K+K_a}{2}\rfloor} \tag{26}$$

where $\lceil x\rceil$ and $\lfloor x\rfloor$ are the smallest integer larger than or equal to x and the largest integer smaller than or equal to x, respectively, and C_0 is a constant given after (25)*.*

Proof: See Appendix B. ■

Using the above lemma, the following main result, which states that the extended measurement matrix $\mathbf{\Phi}_e$ in (24) satisfies the RIP, can be also proved.

Theorem 2. *Let $\mathbf{\Phi}_e$ be formed as in* (24) *and let the elements of $\mathbf{\Phi}$ be i.i.d. zero mean Gaussian variables with variance $1/N$. If min$\{K, K_a+M-1\}\leq\lceil(K+K_a)/2\rceil$, then for any $0<\delta_S<1$, there exist constants C_3 and C_4, which depend only on δ_S, such that for $S\leq C_3\lfloor(K+K_a)/2\rfloor/\log(N/S)$ the inequality* (2) *holds for all S-sparse vectors with probability that satisfies the following inequality*

$$Pr\{\mathbf{\Phi}_e \text{ satisfies RIP}\} \geq 1-4e^{-C_4\lfloor(K+K_a)/2\rfloor} \tag{27}$$

where $C_4=C_0-C_3\left[1+\left(1+\log\left(12/\delta_S\right)\right)/\log\left(N/S\right)\right]$ and C_3 is small enough that guarantees that C_4 is positive.

Proof: See Appendix C. ■

Let us consider now the general case when the number of additional samples K_a is larger than the number of BMIs K, i.e., $K_a > K$, $K_e > 2K$, and the extended measurement matrix is given by (23). Note that while proving Lemma 1 for the special case of Example 3, we were able to split the rows of $\mathbf{\Phi}_e$ into two sets each consisting of independent entries. In the general case, some of the entries of the original measurement matrix appear more than twice in the extended measurement matrix $\mathbf{\Phi}_e$, and it is no longer possible to split the rows of $\mathbf{\Phi}_e$ into only two sets with independent entries. Due to the way that the additional samples are built, the samples $y_{lK+1}, y_{lK+2}, \ldots, y_{(l+1)K}$ obtained based on the permuted matrix $\boldsymbol{Y}^{\mathcal{P}^{(l)}}$, i.e., the l-th set of additional samples, are uncorrelated with each other, but they are correlated with every other set of samples based on the original matrix $\boldsymbol{Y}$ and the permuted matrices $\boldsymbol{Y}^{\mathcal{P}^{(i)}}$, $\forall i,\ i \neq l$. Thus, the following principle can be used while partitioning the rows of $\mathbf{\Phi}_e$ into the sets with independent entries. First, the rows corresponding to the original samples form a single set with independent entries, then the rows corresponding to the first set of additional samples based on the matrix $\boldsymbol{Y}^{\mathcal{P}^{(1)}}$ form another set and so on. Then the number of such sets is $n_p = \lceil K_e/K \rceil$, while the size of each set is

$$K_i = \begin{cases} K, & 1 \leq i < \lceil \frac{K_e}{K} \rceil - 1 \\ K_e - (\lceil \frac{K_e}{K} \rceil - 1)K, & i = \lceil \frac{K_e}{K} \rceil \end{cases} \tag{28}$$

The extended measurement matrix (23) can be rewritten as

$$\mathbf{\Phi}_e = \left((\mathbf{\Phi}_e)_1^T, (\mathbf{\Phi}_e)_2^T, \ldots, (\mathbf{\Phi}_e)_{n_p}^T \right)^T \tag{29}$$

where $(\mathbf{\Phi}_e)_i$ is the i-th partition of $\mathbf{\Phi}_e$ of size given by (28). Then the general form of Lemma 1 is as follows.

Lemma 2. *Let the elements of the measurement matrix $\mathbf{\Phi}$ be i.i.d. zero mean Gaussian variables with variance $1/N$, $\mathbf{\Phi}_e$ be the extended measurement matrix* (23)*, and $\mathcal{T} \subset \{1, \ldots, N\}$ of size S. Let also $K_a > K$ and $n_p = \lceil K_e/K \rceil$. Then, for any $0 < \delta_S < 1$, the following inequality holds*

$$Pr\{(\mathbf{\Phi}_e)_{\mathcal{T}} \text{ satisfies (2)}\} \geq 1 - 2(n_p - 1)\left(12/\delta_S\right)^S \left(e^{-C_0 K}\right) - 2\left(12/\delta_S\right)^S \left(e^{-C_0 K_{n_p}}\right) \tag{30}$$

where $K_{n_p} = K_e - \left(\lceil \frac{K_e}{K} \rceil - 1\right) K$ and C_0 is a constant given after (25)*.*

Proof: See Appendix D. ■

Lemma 2 is needed to prove that the extended measurement matrix (29) satisfies the RIP. Therefore, the general version of Theorem 2 is as follows.

Theorem 3. *Let the elements of $\mathbf{\Phi}$ be i.i.d. zero mean Gaussian variables with variance $1/N$ and $\mathbf{\Phi}_e$ be formed as in* (23)*. If $K_a > K$, then for any $0 < \delta_S < 1$, there exist constants C_3, C_4 and C_4', such that for $S \leq C_3 K_{n_p}/\log(N/S)$ the inequality* (2) *holds for all S-sparse vectors with probability that satisfies the following inequality*

$$Pr\{\mathbf{\Phi}_e \text{ satisfies RIP}\} \geq 1 - 2(n_p - 1)e^{-C_4' K} - 2e^{-C_4 K_{n_p}} \tag{31}$$

where $C_4' = C_0 - (C_3 K_{n_p}/K) \times [1 + (1 + \log(12/\delta_S))/\log(N/S)]$, C_4 is given after (27)*, and C_3 is small enough to guarantee that C_4 and C_4' are both positive.*

Proof: See Appendix E. ■

When splitting the rows of $\mathbf{\Phi}_e$ in a number of sets as described before Lemma 2 it may happen that the last subset $(\mathbf{\Phi}_e)_{n_p}$ has the smallest size K_{n_p}. As a result, the dominant term in (31) will likely be the term $2e^{-C_4 K_{n_p}}$. Moreover, it may lead to a more stringent sparsity condition, that is, $S \leq C_3 K_{n_p}/\log(N/S)$. To improve the lower bound in (31), we can move some of the rows from $(\mathbf{\Phi}_e)_{n_p-1}$ to $(\mathbf{\Phi}_e)_{n_p}$ in order to make the last two partitions of almost the same size. Then the requirement on the sparsity level will become $S \leq C_3 K'/\log(N/S)$ where $K' = \lfloor (K + K_{n_p})/2 \rfloor$. Therefore, the lower bound on the probability calculated in (31) improves.

V. Performance Analysis of the recovery

In this section, we aim at answering the question whether signal recovery also improves if the proposed segmented AIC method, i.e., the extended measurement matrix $\mathbf{\Phi}_e$ (23), is used instead of the original matrix $\mathbf{\Phi}$. The study is performed based on the empirical risk minimization method for signal recovery from noisy random projections [4]. As mentioned in Section II, the LASSO method can be viewed as one of the possible implementations of the empirical risk minimization method.

We first consider the special case of Example 3 when the extended measurement matrix is given by (24). Let the entries of the measurement matrix $\mathbf{\Phi}$ be selected with equal probability as $\pm 1/\sqrt{N}$, i.e., be i.i.d. Bernoulli distributed with variance $1/N$. This assumption is the same as in [4] and it is used here in order to shorten our derivations by only emphasizing the differences caused by our construction of matrix $\mathbf{\Phi}_e$, where some rows are correlated to each other, as compared to the case analyzed in [4], where the measurement matrix consists of all i.i.d. entries. Note that our results can be easily applied to the case of Gaussian distributed entries of $\mathbf{\Phi}$ by only changing the moments of Bernoulli distribution to the moments of Gaussian distribution.

Let $r(\hat{\boldsymbol{f}}, \boldsymbol{f}) \triangleq r(\hat{\boldsymbol{f}}) - r(\boldsymbol{f})$ be the "excess risk" between the candidate reconstruction $\hat{\boldsymbol{f}}$ of the signal sampled using the extended measurement matrix $\boldsymbol{\Phi}_e$ and the actual signal $\boldsymbol{f}$, and $\hat{r}(\hat{\boldsymbol{f}}, \boldsymbol{f}) \triangleq \hat{r}(\hat{\boldsymbol{f}}) - \hat{r}(\boldsymbol{f})$ be the "empirical excess risk" between the candidate signal reconstruction and the actual signal, where $r(\hat{\boldsymbol{f}})$ and $\hat{r}(\hat{\boldsymbol{f}})$ are defined in (6). Then the difference between the "excess risk" and the "empirical excess risk" can be found as

$$r(\hat{\boldsymbol{f}}, \boldsymbol{f}) - \hat{r}(\hat{\boldsymbol{f}}, \boldsymbol{f}) = \frac{1}{K_e} \sum_{j=1}^{K_e} (U_j - E[U_j]) \tag{32}$$

where $U_j \triangleq (y_j - \boldsymbol{\phi}_j \boldsymbol{f})^2 - (y_j - \boldsymbol{\phi}_j \hat{\boldsymbol{f}})^2$.

The mean-square error (MSE) between the candidate reconstruction and the actual signal can be expressed as [24]

$$\text{MSE} \triangleq E\left\{\|\boldsymbol{g}\|^2\right\} = N r(\hat{\boldsymbol{f}}, \boldsymbol{f}) \tag{33}$$

where $\boldsymbol{g} \triangleq \hat{\boldsymbol{f}} - \boldsymbol{f}$. Therefore, if we know an upper bound on the right-hand side of (32), denoted hereafter as U, we can immediately find an upper bound on the MSE in the form $\text{MSE} \leq N\hat{r}(\hat{\boldsymbol{f}}, \boldsymbol{f}) + NU$. In other words, to find the candidate reconstruction $\hat{\boldsymbol{f}}$ one can minimize $\hat{r}(\hat{\boldsymbol{f}}, \boldsymbol{f}) + U$, that will also result in a bound on the MSE as in (8).

The Craig-Bernstein inequality [4], [26] can be used in order to find an upper bound U on the right-hand side of (32). In the notations of our paper, this inequality states that the probability of the following event

$$\frac{1}{K_e} \sum_{j=1}^{K_e} (U_j - E\{U_j\}) \leq \frac{\log\left(\frac{1}{\delta}\right)}{K_e \epsilon} + \frac{\epsilon \, \text{var}\left\{\sum_{j=1}^{K_e} U_j\right\}}{2K_e(1-\zeta)} \tag{34}$$

is greater than or equal to $1 - \delta$ for $0 < \epsilon h \leq \zeta < 1$, if the random variables U_j satisfy the following moment condition for some $h > 0$ and all $k \geq 2$

$$E\left\{|U_j - E\{U_j\}|^k\right\} \leq \frac{k! \, \text{var}\{U_j\} \, h^{k-2}}{2}. \tag{35}$$

The second term in the right-hand side of (34) contains the variance $\text{var}\left\{\sum_{j=1}^{K_e} U_j\right\}$, which we need to calculate or at least find an upper bound on it.

In the case of the extended measurement matrix, the random variables U_j, $j = 1, \ldots, K_e$ all satisfy the moment condition for the Craig-Bernstein inequality [26] with the same coefficient $h = 16B^2 e + 8\sqrt{2}B\sigma$, where σ^2 is the variance of the Gaussian noise.[5] Moreover, it is easy to show that the following bound

[5]The derivation of the coefficient h coincides with a similar derivation in [4], and therefore, is omitted.

on the variance of U_j is valid for the extended measurement matrix[6]

$$\text{var}\{U_j\} \leq \left(2\frac{\|\boldsymbol{g}\|^2}{N} + 4\sigma^2\right)\frac{\|\boldsymbol{g}\|^2}{N} \leq \left(8B^2 + 4\sigma^2\right) r(\hat{\boldsymbol{f}}, \boldsymbol{f}). \tag{36}$$

However, unlike [4], in the case of the extended measurement matrix, the variables U_j are not independent from each other. Thus, we can not simply replace the term var $\left\{\sum_{j=1}^{K_e} U_j\right\}$ with the sum of the variances for $U_j,\ j = 1, \ldots, K_e$. Using the definition of the variance, we can write that

$$\begin{aligned}
\text{var}\left\{\sum_{j=1}^{K_e} U_j\right\} &\triangleq E\left\{\left(\sum_{j=1}^{K_e} U_j\right)^2\right\} - \left(E\left\{\sum_{j=1}^{K_e} U_j\right\}\right)^2 \\
&= \sum_{j=1}^{K_e} E\{U_j^2\} + 2\sum_{i=1}^{K_e-1}\sum_{j=i+1}^{K_e} E\{U_iU_j\} - K_e^2\left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 \\
&= \sum_{j=1}^{K_e}\left(E\{U_j^2\} - \left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2\right) + 2\sum_{i=1}^{K_e-1}\sum_{j=i+1}^{K_e}\left(E\{U_iU_j\} - \left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2\right) \\
&= \sum_{j=1}^{K_e}\text{var}\{U_j\} + 2\sum_{i=1}^{K_e-1}\sum_{j=i+1}^{K_e}\left(E\{U_iU_j\} - \left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2\right)
\end{aligned} \tag{37}$$

where the upper bound on var$\{U_j\}$ is given by (36). Using the fact that the random noise components w_i and w_j are independent from $\boldsymbol{\phi}_i\boldsymbol{g}$ and $\boldsymbol{\phi}_j\boldsymbol{g}$ (see the noisy model (4)), respectively, $E\{U_iU_j\}$ can be expressed as

$$\begin{aligned}
E\{U_iU_j\} &= E\big\{[2w_i\boldsymbol{\phi}_i\boldsymbol{g} - (\boldsymbol{\phi}_i\boldsymbol{g})^2][2w_j\boldsymbol{\phi}_j\boldsymbol{g} - (\boldsymbol{\phi}_j\boldsymbol{g})^2]\big\} \\
&= 4E\big\{w_iw_j\big\}E\big\{\boldsymbol{\phi}_i\boldsymbol{g}\boldsymbol{\phi}_j\boldsymbol{g}\big\} - 2E\big\{w_i\big\}E\big\{\boldsymbol{\phi}_i\boldsymbol{g}(\boldsymbol{\phi}_j\boldsymbol{g})^2\big\} \\
&\quad - 2E\big\{w_j\big\}E\big\{\boldsymbol{\phi}_j\boldsymbol{g}(\boldsymbol{\phi}_i\boldsymbol{g})^2\big\} + E\big\{(\boldsymbol{\phi}_i\boldsymbol{g})^2(\boldsymbol{\phi}_j\boldsymbol{g})^2\big\}.
\end{aligned} \tag{38}$$

The latter expression can be further simplified using the fact that $E\{w_i\} = E\{w_j\} = 0$. Thus, we obtain that

$$E\{U_iU_j\} = 4E\big\{w_iw_j\big\}E\big\{(\boldsymbol{\phi}_i\boldsymbol{g})(\boldsymbol{\phi}_j\boldsymbol{g})\big\} + E\big\{(\boldsymbol{\phi}_i\boldsymbol{g})^2(\boldsymbol{\phi}_j\boldsymbol{g})^2\big\}. \tag{39}$$

It is easy to verify that if $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ are independent, then $E(U_iU_j) = E\left\{(\boldsymbol{\phi}_i\boldsymbol{g})^2\right\}E\left\{(\boldsymbol{\phi}_j\boldsymbol{g})^2\right\} = \left(\|\boldsymbol{g}\|^2/\ N\right)^2$ which indeed coincides with [4]. However, in our case, $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ may depend on each other. If they indeed depend on each other, they have $L = N/M$ common entries, while the rest of the entries are independent. In addition, the additive noise terms w_i and w_j are no longer independent

[6]This bound also coincides with a similar one in [4]

random variables as well and, thus, $E\{w_i w_j\} = \sigma^2/M$. Without loss of generality, let the first L entries of $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ be the same, that is,

$$\boldsymbol{\phi}_i \boldsymbol{g} = \overbrace{g_1 a_1 + \ldots + g_L a_L}^{A} + \overbrace{g_{L+1}\phi_{i,L+1} + \ldots + g_N \phi_{i,N}}^{P_i} \tag{40}$$

$$\boldsymbol{\phi}_j \boldsymbol{g} = \overbrace{g_1 a_1 + \ldots + g_L a_L}^{A} + \overbrace{g_{L+1}\phi_{j,L+1} + \ldots + g_N \phi_{j,N}}^{P_j} \tag{41}$$

with $a_1, ..., a_L$ being the common part between $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$.

Let $\boldsymbol{g}_A$ be a sub-vector of $\boldsymbol{g}$ containing the L elements of $\boldsymbol{g}$ corresponding to the common part between $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$, and $\boldsymbol{g}_{A'}$ be the sub-vector comprising the rest of the elements. Then using the fact that A, P_i, and P_j are all zero mean independent random variables, we can express $E\{(\boldsymbol{\phi}_i \boldsymbol{g})(\boldsymbol{\phi}_j \boldsymbol{g})\}$ from the first term on the right-hand side of (39) as

$$\begin{aligned} E\{(\boldsymbol{\phi}_i \boldsymbol{g})(\boldsymbol{\phi}_j \boldsymbol{g})\} &= E\{(A + P_i)(A + P_j)\} = E\{A^2\} + E\{AP_i\} + E\{AP_j\} + E\{P_i P_j\} \\ &= E\{A^2\} = \frac{\left(\sum_{k=1}^{L} g_k^2\right)^2}{N} = \frac{\|\boldsymbol{g}_A\|^2}{N}. \end{aligned} \tag{42}$$

Similar, the second term on the right-hand side of (39) can be expressed as

$$E\left\{(\boldsymbol{\phi}_i \boldsymbol{g})^2 (\boldsymbol{\phi}_j \boldsymbol{g})^2\right\} = E\left\{(A^2 + P_i^2 + 2AP_i)(A^2 + P_j^2 + 2AP_j)\right\}. \tag{43}$$

Using the facts that $4E\{w_i w_j\} = 4\sigma^2/M$, $E\{A^2\} = \|\boldsymbol{g}_A\|^2/N$, and $E\{P_i^2\} = \|\boldsymbol{g}_{A'}\|^2/N$, the expression (43) can be further rewritten as

$$\begin{aligned} E\left\{(\boldsymbol{\phi}_i \boldsymbol{g})^2 (\boldsymbol{\phi}_j \boldsymbol{g})^2\right\} &= E\left\{A^4 + A^2 P_i^2 + A^2 P_j^2 + P_i^2 P_j^2\right\} = E\{A^4\} + 2\frac{\|\boldsymbol{g}_A\|^2}{N} \cdot \frac{\|\boldsymbol{g}_{A'}\|^2}{N} + \left(\frac{\|\boldsymbol{g}_{A'}\|^2}{N}\right)^2 \\ &= E\{A^4\} + \left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 - \left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2. \end{aligned} \tag{44}$$

Substituting (42) and (44) into (39), we obtain that

$$E\{U_i U_j\} = \frac{4\sigma^2}{M} \cdot \frac{\|\boldsymbol{g}_A\|^2}{N} + E\{A^4\} + \left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 - \left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2. \tag{45}$$

Moreover, substituting (45) into (37), we find that

$$\text{var}\left\{\sum_{j=1}^{K_e} U_j\right\} = \sum_{j=1}^{K_e} \text{var}\{U_j\} + 2 \sum_{\boldsymbol{\phi}_i, \boldsymbol{\phi}_j \text{dependent}} \left(E\{A^4\} - \left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2 + \frac{4\sigma^2}{M} \cdot \frac{\|\boldsymbol{g}_A\|^2}{N}\right). \tag{46}$$

Using the fact that the extended measurement matrix is constructed such that the waveforms $\boldsymbol{\phi}_i$, $i = K+1, \ldots, K_e$ are built upon M rows of the original matrix and also using the inequality[7] $E\{A^4\} -$

[7]We skip the derivation of this inequality since it is relatively well known and can be found, for example, in [4, p. 4039].

$\left(\|\boldsymbol{g}_A\|^2/N\right)^2 \leq 2\left(\|\boldsymbol{g}_A\|^2/N\right)^2$ for all these M rows, we obtain for every $\boldsymbol{\phi}_i$, $i = K+1, \ldots, K_e$ that

$$\sum_{k=1}^{M}\left(E\{A^4\} - \left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right) \leq \sum_{k=1}^{M}\left(2\left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right) \tag{47}$$

where $\boldsymbol{g}_A$ corresponds to the first L entries of $\boldsymbol{g}$ for $k=1$, to the entries from $L+1$ to $2L$ for $k=2$ and so on. Applying also the triangle inequality, we find that

$$\sum_{k=1}^{M}\left(2\left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right) \leq 2\left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}\|^2}{N}. \tag{48}$$

Combining (47) and (48) and using the fact that there are K_a additional rows in the extended measurement matrix, we obtain that

$$2\sum_{\boldsymbol{\phi}_i,\boldsymbol{\phi}_j \text{ dependent}}\left(E\{A^4\} - \left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right) \leq 4K_a\left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 + \frac{8\sigma^2K_a}{M}\cdot\frac{\|\boldsymbol{g}\|^2}{N}. \tag{49}$$

Noticing that $\|\boldsymbol{g}\|^2/N = r(\hat{\boldsymbol{f}},\boldsymbol{f})$ and $\|\boldsymbol{g}\|^2 \leq 4NB^2$, the right-hand side of the inequality (49) can be further upper bounded as

$$4K_a\left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 + \frac{8\sigma^2K_a}{M}\cdot\frac{\|\boldsymbol{g}\|^2}{N} \leq 16K_aB^2\, r(\hat{\boldsymbol{f}},\boldsymbol{f}) + \frac{8\sigma^2K_a}{M}\, r(\hat{\boldsymbol{f}},\boldsymbol{f}). \tag{50}$$

Using the upper bound (50) for the second term in (46) and the upper bound (36) for the first term in (46), we finally can upper bound the var $\left\{\sum_{j=1}^{K_e} U_j\right\}$ as

$$\text{var}\left\{\sum_{j=1}^{K_e} U_j\right\} \leq K_e\left(8B^2\left(1+\frac{2K_a}{K_e}\right) + 4\sigma^2\left(1+\frac{2K_a}{MK_e}\right)\right) r(\hat{\boldsymbol{f}},\boldsymbol{f}). \tag{51}$$

Therefore, based on the Craig-Bernstein inequality, the probability that for a given candidate signal $\hat{\boldsymbol{f}}$ the following inequality holds

$$r(\hat{\boldsymbol{f}},\boldsymbol{f}) - \hat{r}(\hat{\boldsymbol{f}},\boldsymbol{f}) \leq \frac{\log(\frac{1}{\delta})}{K_e\epsilon} + \frac{\left(8B^2\left(1+\frac{2K_a}{K_e}\right) + 4\sigma^2\left(1+\frac{2K_a}{MK_e}\right)\right) r(\hat{\boldsymbol{f}},\boldsymbol{f})\,\epsilon}{2(1-\zeta)} \tag{52}$$

is greater than or equal to $1-\delta$.

Let $c(\hat{\boldsymbol{f}})$ be chosen such that the Kraft inequality $\sum_{\hat{\boldsymbol{f}}\in\mathcal{F}(B)} 2^{c(\hat{\boldsymbol{f}})} \leq 1$ is satisfied (see also [4]), and let $\delta(\hat{\boldsymbol{f}}) = 2^{-c(\hat{\boldsymbol{f}})}\,\delta$. Applying the union bound to (52), it can be shown that for all $\hat{\boldsymbol{f}} \in \mathcal{F}(B)$ and for all $\delta > 0$, the following inequality holds with probability of at least $1-\delta$

$$r(\hat{\boldsymbol{f}},\boldsymbol{f}) - \hat{r}(\hat{\boldsymbol{f}},\boldsymbol{f}) \leq \frac{c(\hat{\boldsymbol{f}})\log 2 + \log(\frac{1}{\delta})}{K_e\,\epsilon} + \frac{\left(8B^2\left(1+\frac{2K_a}{K_e}\right) + 4\sigma^2\left(1+\frac{2K_a}{MK_e}\right)\right) r(\hat{\boldsymbol{f}},\boldsymbol{f})\,\epsilon}{2(1-\zeta)}. \tag{53}$$

Finally, setting $\zeta = \epsilon\, h$ and

$$a = \frac{\left(8B^2\left(1+\frac{2K_a}{K_e}\right)+4\sigma^2\left(1+\frac{2K_a}{MK_e}\right)\right)\epsilon}{2(1-\zeta)} \tag{54}$$

$$\epsilon < \frac{1}{\left(4\left(1+\frac{2K_a}{K_e}\right)+16e\right)B^2+8\sqrt{B}\sigma+2\sigma^2\left(1+\frac{2K_a}{MK_e}\right)} \tag{55}$$

where $0 < \epsilon\, h \leq \zeta < 1$ as required by the Craig-Bernstein inequality, the following inequality holds with probability of at least $1-\delta$ for all $\hat{\boldsymbol{f}} \in \mathcal{F}(B)$

$$(1-a)r(\hat{\boldsymbol{f}},\boldsymbol{f}) \leq \hat{r}(\hat{\boldsymbol{f}},\boldsymbol{f}) + \frac{c(\hat{\boldsymbol{f}})\log 2 + \log(\frac{1}{\delta})}{K_e\,\epsilon}. \tag{56}$$

The following result on the recovery performance of the empirical risk minimization method is in order.

Theorem 4. *Let ϵ be chosen as*

$$\epsilon = \frac{1}{\left(60\,(B+\sigma)^2\right)} \tag{57}$$

which satisfies the inequality (55)*, then the signal reconstruction $\hat{\boldsymbol{f}}_{K_e}$ given by*

$$\hat{\boldsymbol{f}}_{K_e} = \arg\min_{\hat{\boldsymbol{f}}\in\mathcal{F}(B)}\left\{\hat{r}(\hat{\boldsymbol{f}}) + \frac{c(\hat{\boldsymbol{f}})\log 2}{\epsilon K_e}\right\} \tag{58}$$

satisfies the following inequality

$$E\left\{\frac{\|\hat{\boldsymbol{f}}_{K_e}-\boldsymbol{f}\|^2}{N}\right\} \leq C_{1e}\min_{\hat{\boldsymbol{f}}\in\mathcal{F}(B)}\left\{\frac{\|\hat{\boldsymbol{f}}-\boldsymbol{f}\|^2}{N} + \frac{c(\hat{\boldsymbol{f}})\log 2+4}{\epsilon K_e}\right\} \tag{59}$$

where C_{1e} is the constant given as

$$C_{1e} = \frac{1+a}{1-a}, \quad a = \frac{2\left(1+\frac{2K_a}{K_e}\right)\left(\frac{B}{\sigma}\right)^2+\left(1+\frac{2K_a}{MK_e}\right)}{(30-8e)\left(\frac{B}{\sigma}\right)^2+\left(60-4\sqrt{2}\right)\left(\frac{B}{\sigma}\right)+30} \tag{60}$$

with a obtained from (54) *for the specific choice of ϵ in* (57)*.*

Proof: The proof follows the exact steps of the proof of the related result for the uncorrelated case [4, p. 4039–4040] with the exception of using, in our correlated case, the above calculated values for ϵ (57) and a (60). ■

Example 4: Let one set of samples be obtained based on the measurement matrix $\boldsymbol{\Phi}_e$ with $K_a = K$, $K_e = 2K$, and $M = 8$, and let another set of samples be obtained using a $2K \times N$ measurement matrix with all i.i.d. (Bernoulli) elements. Let also ϵ be selected as (57). Then the MSE error bounds for these two cases differ from each other only by a constant factor given for the former case by C_{1e} in (60) and in the latter case by C_1 (see (8) and the row after). Considering the two limiting cases when

$B/\sigma \to 0$ and $B/\sigma \to \infty$, the intervals of change for the corresponding coefficients can be obtained as $1.08 \leq C_{1e} \leq 2.88$ and $1.06 \leq C_1 \leq 1.63$, respectively.

The following result on the achievable recovery performance for a sparse or compressible signal sampled based on the extended measurement matrix $\mathbf{\Phi}_e$ is also of great interest.

Theorem 5. *For a sparse signal $\boldsymbol{f} \in \mathcal{F}_s(B,S) = \{\boldsymbol{f} : \|\boldsymbol{f}\|^2 \leq NB^2, \|\boldsymbol{f}\|_{l_0} \leq S\}$ and corresponding reconstructed signal $\hat{\boldsymbol{f}}_{K_e}$ obtained according to* (58)*, there exists a constant $C'_{2e} = C'_{2e}(B,\sigma) > 0$, such that*

$$\sup_{\boldsymbol{f} \in \mathcal{F}_s(B,S)} E\left\{\frac{\|\hat{\boldsymbol{f}}_{K_e} - \boldsymbol{f}\|^2}{N}\right\} \leq C_{1e}C'_{2e}\left(\frac{K_e}{S\log N}\right)^{-1}. \tag{61}$$

Similar, for a compressible signal $\boldsymbol{f} \in \mathcal{F}_c(B,\alpha,C_A) = \{\boldsymbol{f} : \|\boldsymbol{f}\|^2 \leq NB^2, \|\boldsymbol{f}^{(m)} - \boldsymbol{f}\|^2 \leq NC_A m^{-2\alpha}\}$ and corresponding reconstructed signal $\hat{\boldsymbol{f}}_{K_e}$ obtained according to (58)*, there exists a constant $C_{2e} = C_{2e}(B,\sigma,C_A) > 0$, such that*

$$\sup_{\boldsymbol{f} \in \mathcal{F}_c(B,\alpha,C_A)} E\left\{\frac{\|\hat{\boldsymbol{f}}_{K_e} - \boldsymbol{f}\|^2}{N}\right\} \leq C_{1e}C_{2e}\left(\frac{K_e}{\log N}\right)^{-2\alpha/(2\alpha+1)}. \tag{62}$$

Proof: The proof follows the exact steps of the proofs of the related results for the uncorrelated case [4, p. 4040–4041] with the exception of using, in our correlated case, the above calculated values for ϵ (57) and a (60). ■

Example 5: Let one set of samples be obtained based on the extended measurement matrix $\mathbf{\Phi}_e$ with $K_a = K$, $K_e = 2K$, and $M = 8$ and let another set of samples be obtained using the $K \times N$ measurement matrix with all i.i.d. (Bernoulli) elements. The error bounds corresponding to the case of K uncorrelated samples of [4] and our case of K_e correlated samples are (10) and (61), respectively. The comparison between these two error bounds boils down in this example to comparing $2C_1C'_2$ and $C_{1e}C'_{2e}$. Assuming the same ϵ as (57) for both methods, the following holds true $C'_{2e} = C'_2$. Fig. 3 compares C_{1e} and $2C_1$ versus the signal-to-noise ratio (SNR) B^2/σ^2. Since $C_{1e} < 2C_1$ for all values of SNR, the quality of the signal recovery, i.e., the corresponding MSE, for the case of $2K \times N$ extended measurement matrix is expected to be better than the quality of the signal recovery for the case of $K \times N$ measurement matrix of all i.i.d. entries.

The above results can be easily generalized for the case when $K_a > K$. Indeed, we only need to recalculate $\text{var}\left\{\sum_{j=1}^{K_e} U_j\right\}$ for $K_a > 2K$. The only difference with the previous case of $K_a \leq K$ is the increased number of pairs of dependent rows in the extended measurement matrix $\mathbf{\Phi}_e$, which has a larger size now. The latter affects only the second term in (46). In particular, every row in $\mathbf{\Phi}^{\mathcal{P}^{(1)}}$ depends on M rows of the original measurement matrix $\mathbf{\Phi}$. Moreover, the term $\sum_{i=1}^{2K-1}\sum_{j=i+1}^{2K} E\{U_iU_j\}$ over

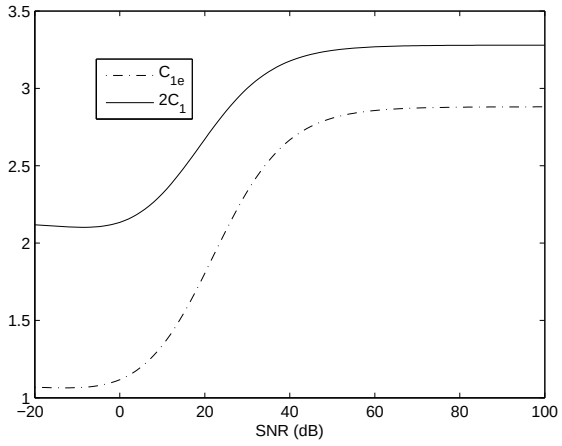


Fig. 3. C_{1e} and $2C_1$ versus SNR.

all these M rows is bounded as in (48). Then considering all KM pairs of dependent rows from $\mathbf{\Phi}$ and $\mathbf{\Phi}^{\mathcal{P}^{(1)}}$, we have

$$2\sum_{\boldsymbol{\phi}_i,\boldsymbol{\phi}_j\text{dependent}}\left(E\{A^4\}-\left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2+\frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right)\leq 4K\left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2+\frac{8\sigma^2K}{M}\cdot\frac{\|\boldsymbol{g}\|^2}{N}. \tag{63}$$

Similar, every row of $\mathbf{\Phi}^{\mathcal{P}^{(2)}}$ depends on M rows of $\mathbf{\Phi}^{\mathcal{P}^{(1)}}$ and M rows of $\mathbf{\Phi}$. Considering all these $2KM$ pairs of dependent rows, we have

$$2\sum_{\boldsymbol{\phi}_i,\boldsymbol{\phi}_j\text{dependent}}\left(E\{A^4\}-\left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2+\frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right)\leq 4(2K)\left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2+\frac{8\sigma^2(2K)}{M}\cdot\frac{\|\boldsymbol{g}\|^2}{N}. \tag{64}$$

Finally, the number of rows in the last matrix $(\mathbf{\Phi}_e)_{n_p}$ is K_{n_p} (see (28) and (29)). Every row of $(\mathbf{\Phi}_e)_{n_p}$ depends on M rows of each of the previous n_p-1 matrices $\mathbf{\Phi}^{\mathcal{P}^{(i)}}$, $i=1,\ldots,n_p-1$. Considering all $(n_p-1)K_{n_p}M$ pairs of dependent rows, we have

$$2\sum_{\boldsymbol{\phi}_i,\boldsymbol{\phi}_j\text{dependent}}\left(E\{A^4\}-\left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2+\frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right)\leq 4(n_p-1)K_{n_p}\left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2+\frac{8\sigma^2(n_p-1)K_{n_p}}{M}\cdot\frac{\|\boldsymbol{g}\|^2}{N}. \tag{65}$$

Based on the equations (37) and (63)–(65) we can find the following bound

$$\text{var}\left\{\sum_{j=1}^{K_e}U_j\right\}\leq K_e\left(8B^2\left(1+\frac{D}{K_e}\right)+4\sigma^2\left(1+\frac{D}{MK_e}\right)\right)r(\hat{\boldsymbol{f}},\boldsymbol{f}) \tag{66}$$

where $D=2K\sum_{i=1}^{n_p-2}i+2K_{n_p}(n_p-1)$. Note that in the case that $K_e=n_pK$, we have $D/K_e=n_p-1$.

Therefore, it can be shown for the general extended matrix (23) that the inequality (56) holds with the following values of a and ϵ:

$$a = \frac{\left(8B^2\left(1+\frac{D}{K_e}\right)+4\sigma^2\left(1+\frac{D}{MK_e}\right)\right)\epsilon}{2(1-\zeta)} \tag{67}$$

$$\epsilon < \frac{1}{\left(4\left(1+\frac{D}{K_e}\right)+16e\right)B^2+8\sqrt{B}\sigma+2\sigma^2\left(1+\frac{D}{MK_e}\right)} \tag{68}$$

Moreover, the theorems similar to Theorems 4 and 5 follow straightforwardly with the corrections to a and ϵ which are given now by (67) and (68), respectively.

We finally make some remarks on *non-RIP* conditions for l_1-norm-based recovery. Since the extended measurement matrix of the proposed segmented compressed sampling method satisfies the RIP, the results of [23] on recoverability and stability of the l_1-norm minimization straightforwardly apply. A different non-RIP-based approach for studying the recoverability and stability of the l_1-norm minimization, which uses some properties of the null space of the measurement matrix, is used in [27]. Then the non-RIP sufficient condition for recoverability of a sparse signal from its noiseless compressed samples with the algorithm (3) is [27]

$$\sqrt{S} < \min\left\{0.5\frac{\|\boldsymbol{v}\|_{l_1}}{\|\boldsymbol{v}\|_{l_2}} :\ \boldsymbol{v}\in\{\mathcal{N}(\boldsymbol{\Phi})\setminus\{0\}\}\right\} \tag{69}$$

where $\mathcal{N}(\boldsymbol{\Phi})$ denotes the null space of the measurement matrix $\boldsymbol{\Phi}$.

Let us show that the condition (69) is also satisfied for the extended measurement matrix $\boldsymbol{\Phi}_e$. Let $\boldsymbol{d}$ be any vector in the null space of $\boldsymbol{\Phi}_e$, i.e., $\boldsymbol{d}\in\mathcal{N}(\boldsymbol{\Phi}_e)$. Therefore, $[\boldsymbol{\Phi}_e]_i\boldsymbol{d}=0,\ i=1,\ldots,K_e$ where $[\boldsymbol{\Phi}_e]_i$ is the i-th $1\times N$ row-vector of $\boldsymbol{\Phi}_e$. Since the first K rows of $\boldsymbol{\Phi}_e$ are exactly the same as the K rows of $\boldsymbol{\Phi}$, we have $[\boldsymbol{\Phi}]_i\boldsymbol{d}=0,\ i=1,\ldots,K$. Therefore, $\boldsymbol{d}\in\mathcal{N}(\boldsymbol{\Phi})$, and we can conclude that $\mathcal{N}(\boldsymbol{\Phi}_e)\subset\mathcal{N}(\boldsymbol{\Phi})$. Due to this property, we have $\min\{0.5\|\boldsymbol{v}\|_{l_1}/\|\boldsymbol{v}\|_{l_2}:\ \boldsymbol{v}\in\mathcal{N}(\boldsymbol{\Phi})\}\ \leq\ \min\{0.5\|\boldsymbol{v}\|_{l_1}/\|\boldsymbol{v}\|_{l_2}:\ \boldsymbol{v}\in\mathcal{N}(\boldsymbol{\Phi}_e)\}$. Therefore, if the original measurement matrix $\boldsymbol{\Phi}$ satisfies (69), so does the extended measurement matrix $\boldsymbol{\Phi}_e$, and the signal is recoverable from the samples taken by $\boldsymbol{\Phi}_e$.

Moreover, the necessary and sufficient condition for all signals with $\|\boldsymbol{x}\|_{l_0}<S$ to be recoverable from noiseless compressed samples using the l_1-norm minimization (3) is that [27]

$$\|\boldsymbol{v}\|_{l_1} > 2\|\boldsymbol{v}_{\mathcal{T}}\|_{l_1},\ \forall\boldsymbol{v}\in\{\mathcal{N}(\boldsymbol{\Phi})\setminus\{0\}\} \tag{70}$$

where $\mathcal{T}$ is the set of indexes corresponding to the nonzero coefficients of $\boldsymbol{x}$. It is easy to see that since $\mathcal{N}(\boldsymbol{\Phi}_e)\subset\mathcal{N}(\boldsymbol{\Phi})$, the condition (70) also holds for the extended measurement matrix if the original measurement matrix satisfies it.

VI. Simulation Results

Throughout our simulations we use the sparse signal of dimension 128 with only 3 nonzero entries, which are set to ± 1 with equal probabilities. Since the signal is sparse in the time domain, $\boldsymbol{\Psi} = \boldsymbol{I}$. The collected samples are assumed to be noisy, i.e., the model (4) is applied. In all our simulation examples, three different measurement matrices (sampling schemes) are used: (i) the $K \times N$ measurement matrix $\boldsymbol{\Phi}$ with i.i.d. entries referred to as the original measurement matrix; (ii) the extended $K_e \times N$ measurement matrix $\boldsymbol{\Phi}_e$ obtained using the proposed segmented compressed sampling method and referred to as the extended measurement matrix; and (iii) the $K_e \times N$ measurement matrix with all i.i.d entries referred to as the enlarged measurement matrix. This last measurement matrix corresponds to the sampling scheme with K_e independent BMIs in the AIC in Fig. 1. The number of segments in the proposed segmented compressed sampling method M is set to 8. To make sure that the measurement noise for additional samples obtained based on the extended measurement matrix is correlated with the measurement noise of the original samples, the $K \times M$ matrix of noisy sub-samples with the noise variance σ^2/M is first generated. Then the permutations are applied to this matrix and the sub-samples along each row of the original and permuted matrices are added up together to build the noisy samples.

The recovery performance for three aforementioned sampling schemes is measured using the MSE between the recovered and original signals. In all examples, MSE values are computed based on 5000 independent simulation runs for all sampling schemes tested. The SNR is defined as $\|\boldsymbol{\Phi f}\|_{l_2}^2/\|\boldsymbol{w}\|_{l_2}^2$. Approximating $\|\boldsymbol{\Phi f}\|_{l_2}^2$ by $(K'/N)\|\boldsymbol{f}\|_{l_2}^2$, which is valid because of (2), the corresponding noise variance σ^2 can be calculated if SNR is given, and vise versa. Here $K' = K$ for the sampling scheme based on the original measurement matrix, while $K' = K_e$ in the other two schemes. For example, the approximate SNR in dBs can be calculated as $10\log_{10}(3/N\sigma^2)$.

Recovery based on the l_1-norm minimization algorithm: In our first simulation example, the l_1-norm minimization algorithm (5) is used to recover a signal sampled using the three aforementioned sampling schemes. Since $\boldsymbol{\Psi} = \boldsymbol{I}$, then $\boldsymbol{\Phi}' = \boldsymbol{\Phi}$ in (5). The number of BMIs in the sampling device is taken to be $K = 16$, while γ in (5), which is the bound on the root square of the noise energy, is set to $\sqrt{K'}\sigma$. The entries of the original and enlarged measurement matrices are generated as i.i.d. Gaussian distributed random variables with zero mean and variance $1/N$.

Fig. 4 shows the MSEs corresponding to all three aforementioned measurement matrices versus the ratio of the number of additional samples to the number of original samples K_a/K. The results are shown for three different SNR values of 5, 15 and 25 dB. It can be seen from the figure that better recovery

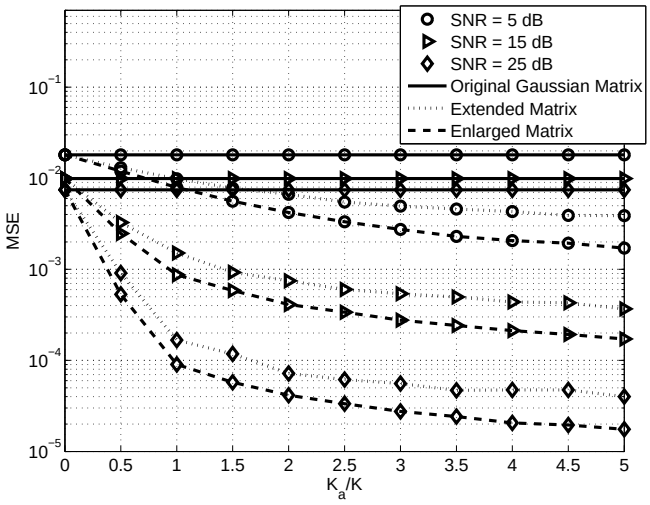


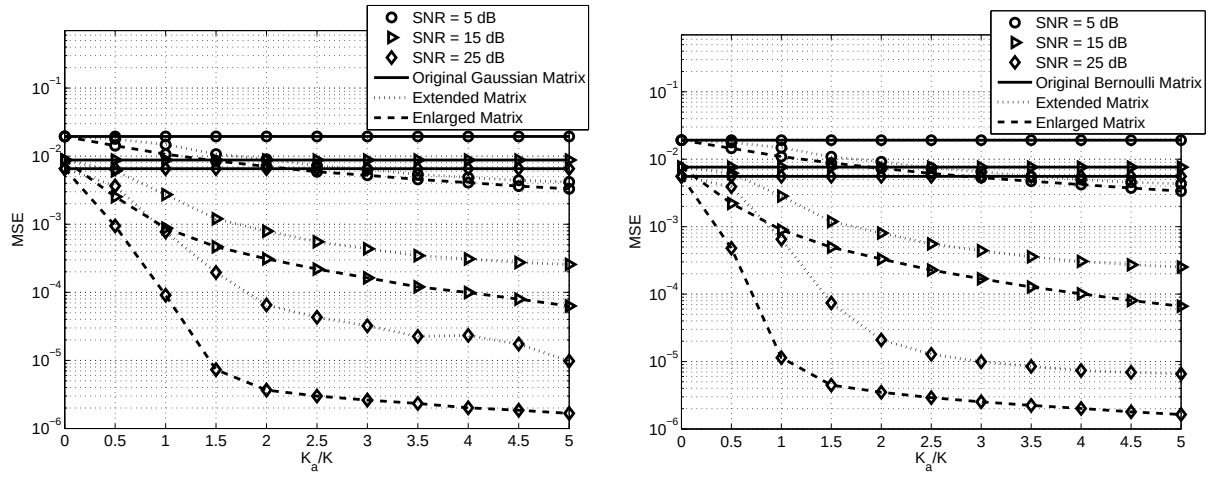
Fig. 4. Recovery based on the l_1-norm minimization algorithm: MSEs versus K_a/K.

quality is achieved by using the extended measurement matrix as compared to the original measurement matrix. The improvements are more significant for high SNRs since the recovery error is proportional to the noise power [23]. As expected, the recovery performance in the case of the extended measurement matrix is not as good as in the case of the enlarged measurement matrix. This difference, however, is small as compared to the performance improvement over the original measurement matrix. Note also that in the case of the enlarged measurement matrix, the AIC in Fig. 1 consists of K_e BMIs, while only K BMIs are required in the case of the extended measurement matrix. Thus, the segmented AIC requires $K_e - K$ less BMIs. For example, the number of such BMIs halves if $K_a/K = 1$. Additionally, it can be seen that the rate of MSE improvement decreases as the number of collected samples increases. The latter can be observed for both the extended and enlarged measurement matrices and for all three values of SNR.

Recovery based on the empirical risk minimization method: In our second simulation example, the empirical risk minimization method is used to recover a signal sampled using the three aforementioned sampling schemes tested with $K = 24$. The minimization problem (7) is solved to obtain a candidate reconstruction $\hat{\boldsymbol{f}}_{K'}$ of the original sparse signal $\boldsymbol{f}$. Considering $\hat{\boldsymbol{f}}_{K'} = \boldsymbol{\Psi}^H \hat{\boldsymbol{x}}_{K'}$, the problem (7) can be rewritten in terms of $\hat{\boldsymbol{x}}_{K'}$ as

$$\hat{\boldsymbol{x}}_{K'} = \arg\min_{\hat{\boldsymbol{x}} \in \mathcal{X}} \left\{ \hat{r}(\boldsymbol{\Psi}^H \hat{\boldsymbol{x}}) + \frac{c(\hat{\boldsymbol{x}}) \log 2}{\epsilon K'} \right\} = \arg\min_{\hat{\boldsymbol{x}} \in \mathcal{X}} \left\{ \|(y) - \boldsymbol{\Phi}\boldsymbol{\Psi}^H \hat{\boldsymbol{x}}\|_{l_2}^2 + \frac{2 \log 2 \log N}{\epsilon} \|\hat{\boldsymbol{x}}\|_{l_0} \right\} \quad (71)$$

and solved using the iterative bound optimization procedure [4]. This procedure uses the threshold



(a) Measurement matrix with Gaussian distributed entries (b) Measurement matrix with Bernoulli distributed entries

Fig. 5. Recovery based on the empirical risk minimization method: MSEs versus K_a/K.

$\sqrt{2\log 2\log N/\lambda\epsilon}$, where λ is the largest eigenvalue of the matrix $\mathbf{\Phi}^T\mathbf{\Phi}$. In our simulations, this threshold is set to 0.035 for the case of the extended measurement matrix and 0.05 for the cases of the original and enlarged measurement matrices. These threshold values are optimized as recommended in [4]. The stopping criterion for the iterative bound optimization procedure is $\|\hat{\boldsymbol{x}}^{(i+1)} - \hat{\boldsymbol{x}}^{(i)}\|_{l_\infty} \leq \theta$, where $\|.\|_{l_\infty}$ is the l_∞ norm and $\hat{\boldsymbol{x}}^{(i)}$ denotes the value of $\hat{\boldsymbol{x}}$ obtained in the i-th iteration. The value $\theta = 0.001$ is selected.

Fig. 5 shows the MSEs obtained based on the empirical risk minimization method for all three measurement matrices versus the ratio K_a/K. The results are shown for three different SNR values of 5, 15 and 25 dB. Two cases are considered: (a) the entries of the original and the enlarged measurement matrices are generated as i.i.d. zero mean Gaussian distributed random variables with variance $1/N$ and (b) the entries of the original and enlarged measurement matrices are generated as i.i.d. zero mean Bernoulli distributed random variables with variance as in case (a). The same conclusions as in the first example can be drawn in this example. Moreover, the results for cases (a) and (b) are also similar. Therefore, the proposed segmented AIC indeed leads to significantly improved signal recovery performance without increasing the number of BMIs.

VII. Conclusion

A new segmented compressed sampling method for AIC has been proposed. According to this method, signal is segmented into M segments and passed through K BMIs of AIC to generate a $K \times M$ matrix of sub-samples. Then, a number of correlated samples larger than the number of BMIs is constructed by

adding up different subsets of sub-samples selected in a specific manner. Due to the inherent structure of the method, the complexity of the sampling device is almost unchanged, while the signal recovery performance is shown to be significantly improved. The complexity increase is only due to the M times higher sampling rate and the necessity to solve a larger size optimization problem at the recovery stage, while the number of BMIs remains the same at the sampling stage. The validity and superiority of the proposed segmented AIC method over the conventional AIC is justified through theoretical analysis of the RIP and the quality of signal recovery. Simulation results also verify the effectiveness and superiority of the proposed segmented AIC method and approve our theoretical studies.

APPENDIX A: PROOF OF THEOREM 1

The total number of possible permutations of $\boldsymbol{z}$ is $K!$. Let $\mathcal{A}$ be the set of permutations $\pi_s,\ s = 1,\ldots,|\mathcal{A}|$ that satisfy the following condition

$$\pi_s(k) \neq \pi_t(k), \quad s \neq t,\ \forall s,t \in \{1,\ldots,|\mathcal{A}|\},\ \forall k \in \{1,\ldots,K\}. \tag{72}$$

It is easy to see that the number of distinct permutations satisfying the condition (72) is K, so $|\mathcal{A}| = K$. It is also straightforward to see that the choice of such K distinct permutations is not unique. As a specific choice, let the elements of $\mathcal{A}$, i.e., the permutations $\pi_s,\ s = 1,\ldots,K$, be

$$\pi_s(k) = ((s+k-2) \bmod K) + 1, \quad s,k = 1,\ldots,K \tag{73}$$

with π_1 being the identity permutation, i.e., the permutations that does not change $\boldsymbol{z}$.

Consider now the matrix $\boldsymbol{Z}$ which consists of M columns $\boldsymbol{z}$. The i-th set of column permutations of matrix $\boldsymbol{Z}$ is $\mathcal{P}^{(i)} = \{\pi_1^{(i)},\ldots,\pi_M^{(i)}\}$ and the corresponding permuted matrix is $\boldsymbol{Z}^{\mathcal{P}^{(i)}}$. Let $\{\pi_1^{(i)},\ldots,\pi_M^{(i)}\}$ be any combination of the K permutations in (73). Then there are K^M possible choices for $\mathcal{P}^{(i)}$. However, not all of these possible choices are permissible by the conditions of the theorem.

Indeed, let the set $\mathcal{P}^{(1)}$ be a combination of permutations from $\mathcal{A}$ that satisfies (18). There are $I-1$ other sets $\mathcal{P}^{(i)},\ i = 2,\ldots,I$ which satisfy both (18) and (19). Gathering all such sets in one set, we obtain the set $\mathcal{P} = \{\mathcal{P}^{(1)},\ldots,\mathcal{P}^{(I)}\}$. Now let $\mathcal{P}^{(I+1)} = [\pi_1^{(I+1)},\ldots,\pi_M^{(I+1)}]$ be one more set of permutations where $\exists \pi_m^{(I+1)},\ m = 1,\ldots,M$ such that $\pi_m^{(I+1)} \notin \mathcal{A}$. An arbitrary k-th row of $\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}$ is $\left([\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}]_{k,1},\ldots,[\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}]_{k,M}\right)$ where $[\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}]_{k,1},\ldots,[\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}]_{k,M} \in \{1,\ldots,K\}$. This exact same row can be found as the first row of one of the permuted matrices $\boldsymbol{Z}^{\mathcal{P}^{(i)}},\ \mathcal{P}^{(i)} \in \mathcal{P}$. Specifically, this is the permuted matrix $\boldsymbol{Z}^{\mathcal{P}^{(i)}}$ that is obtained by applying the permutations $\mathcal{P}^{(i)} = \left\{\pi_{[\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}]_{k,1}},\ldots,\pi_{[\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}]_{k,M}}\right\}$. The permutations $\mathcal{P}^{(i)}$ either has to belong to $\mathcal{P}$ or being crossed out

from $\mathcal{P}$ because of conflicting with some other element $\mathcal{P}^{(l)} \in \mathcal{P}$, $l \neq i$. In both cases, $\mathcal{P}^{(I+1)}$ can not be added to $\mathcal{P}$ because it will contradict the conditions (18) and (19).

Therefore, the set $\mathcal{P}$ can be built using only the permutations from the set $\mathcal{A}$, i.e., the K permutations in (73). Rearranging the rows of $\boldsymbol{Z}^{\mathcal{P}^{(i)}}$ in a certain way, one can force the elements in the first column of $\boldsymbol{Z}^{\mathcal{P}^{(i)}}$ to appear in the original increasing order, i.e., enforce the first column be equivalent to the vector of indexes $\boldsymbol{z}$. It can be done by applying to each permutation in the set $\mathcal{P}^{(i)}$ the inverse permutation $\left(\pi_1^{(i)}\right)^{-1}$, which itself is one of the permutations in (73). Therefore, the set $\mathcal{P}^{(i)} = \{\pi_1^{(i)}, \ldots, \pi_M^{(i)}\}$ can be replaced by the equivalent set $\left\{\left(\pi_1^{(i)}\right)^{-1}\pi_1^{(i)}, \ldots, \left(\pi_1^{(i)}\right)^{-1}\pi_M^{(i)}\right\} = \left\{\pi_1, \ldots, \left(\pi_1^{(i)}\right)^{-1}\pi_M^{(i)}\right\}$, where π_1 is the identity permutation and $\left(\pi_1^{(i)}\right)^{-1}\pi_j^{(i)} \in \mathcal{A}$. Hence, we can consider only the permutations of the form $\mathcal{P}^{(i)} = \{\pi_1, \ldots, \pi_j^{(i)}, \ldots, \pi_M^{(i)}\}$. Since the condition (18) requires that $\pi_2^{(i)}$ should be different from π_1, the only available options for the permutations on the second column of $\boldsymbol{Z}$ are the $K-1$ permutations $\pi_2, \ldots, \pi_K$ in (73). Therefore, I at most equals $K-1$. Note that I can be smaller than $K-1$ if for some $i \in \{1, \ldots, K-1\}$, $K/gcd(i,K) < M$ (also see Example 1 after Theorem 1). Thus, in general $I \leq K-1$.

APPENDIX B: PROOF OF LEMMA 1

Let all the rows of $(\boldsymbol{\Phi}_e)_{\mathcal{T}}$ be partitioned into two sets of sizes (cardinality) as close as possible to each other, where all elements in each set are guaranteed to be statistically independent. In particular, note that the elements of the new K_a rows of $\boldsymbol{\Phi}_e$ are chosen either from the first $K_a + M - 1$ rows of $\boldsymbol{\Phi}$ if $K_a + M - 1 < K$ or from the whole matrix $\boldsymbol{\Phi}$. Therefore, if $K_a + M - 1 < K$, the last $K - K_a - M + 1$ rows of $\boldsymbol{\Phi}$ play no role whatsoever in the process of extending the measurement matrix and they are independent on the rows of $\boldsymbol{\Phi}_1$ in (24). These rows are called unused rows. Thus, one can freely add any number of such unused rows to the set of rows in $\boldsymbol{\Phi}_1$ without disrupting its status of being formed by independent Gaussian variables. Since $\min\{K, K_a + M - 1\} \leq \lceil (K + K_a)/2 \rceil$, there exist at least $\lfloor (K + K_a)/2 \rfloor - K_a$ unused rows which can be added to the set of rows in $\boldsymbol{\Phi}_1$. Such process describes how the rows of $(\boldsymbol{\Phi}_e)_{\mathcal{T}}$ are split into the desired sets $(\boldsymbol{\Phi}_e)_{\mathcal{T}}^1$ and $(\boldsymbol{\Phi}_e)_{\mathcal{T}}^2$ of statistically independent elements. As a result, the first matrix $(\boldsymbol{\Phi}_e)_{\mathcal{T}}^1$ includes the first $\lceil (K + K_a)/2 \rceil$ rows of $(\boldsymbol{\Phi}_e)_{\mathcal{T}}$, while the rest of the rows are included in $(\boldsymbol{\Phi}_e)_{\mathcal{T}}^2$.

Since the elements of the matrices $(\boldsymbol{\Phi}_e)_{\mathcal{T}}^1$ and $(\boldsymbol{\Phi}_e)_{\mathcal{T}}^2$ are i.i.d. Gaussian, they will satisfy (2) with probabilities equal or larger than $1 - 2\left(12/\delta_S\right)^S e^{-C_0 \lceil K_e/2 \rceil}$ and $1 - 2\left(12/\delta_S\right)^S e^{-C_0 \lfloor K_e/2 \rfloor}$, respectively.

Therefore, both matrices $(\mathbf{\Phi}_e)^1_{\mathcal{T}}$ and $(\mathbf{\Phi}_e)^2_{\mathcal{T}}$ satisfy (2) simultaneously with the common probability

$$\Pr\{(\mathbf{\Phi}_e)^i_{\mathcal{T}} \text{ satisfies (2)}\} \geq 1 - 2(12/\delta_S)^S e^{-C_0\lfloor K_e/2\rfloor}, \quad i = 1, 2. \tag{74}$$

Let $K_1' \triangleq \lceil K_e/2 \rceil$ and $K_2' \triangleq \lfloor K_e/2 \rfloor$. Consider the event when both $(\mathbf{\Phi}_e)^1_{\mathcal{T}}$ and $(\mathbf{\Phi}_e)^2_{\mathcal{T}}$ satisfy (2). Then the following inequality hold for any vector $\boldsymbol{c} \in \mathbb{R}^S$:

$$\sum_{i=1}^{2} \frac{K_i'}{N}(1-\delta_S)\|\boldsymbol{c}\|_{l_2}^2 \leq \sum_{i=1}^{2} \|(\mathbf{\Phi}_e)^i_{\mathcal{T}}\boldsymbol{c}\|_{l_2}^2 \leq \sum_{i=1}^{2} \frac{K_i'}{N}(1+\delta_S)\|\boldsymbol{c}\|_{l_2}^2 \tag{75}$$

or, equivalently,

$$\frac{K_e}{N}(1-\delta_S)\|\boldsymbol{c}\|_{l_2}^2 \leq \|(\mathbf{\Phi}_e)_{\mathcal{T}}\boldsymbol{c}\|_{l_2}^2 \leq \frac{K_e}{N}(1+\delta_S)\|\boldsymbol{c}\|_{l_2}^2. \tag{76}$$

Therefore, if both matrices $(\mathbf{\Phi}_e)^1_{\mathcal{T}}$ and $(\mathbf{\Phi}_e)^2_{\mathcal{T}}$ satisfy (2), then the matrix $(\mathbf{\Phi}_e)_{\mathcal{T}}$ also satisfies (2). Moreover, the probability that $(\mathbf{\Phi}_e)_{\mathcal{T}}$ does not satisfy (2) can be found as

$$\begin{aligned}
\Pr\{(\mathbf{\Phi}_e)_{\mathcal{T}} \text{ does not satisfy (2)}\} &\leq \Pr\{(\mathbf{\Phi}_e)^1_{\mathcal{T}} \text{ or } (\mathbf{\Phi}_e)^2_{\mathcal{T}} \text{ does not satisfy (2)}\} \\
&\overset{(a)}{\leq} \sum_{i=1}^{2} \Pr\{(\mathbf{\Phi}_e)^i_{\mathcal{T}} \text{ does not satisfy (2)}\} \\
&\overset{(b)}{\leq} 4\,(12/\delta_S)^S\, e^{-C_0\lfloor K_e/2\rfloor}
\end{aligned} \tag{77}$$

where the inequality (a) follows from the union bounding and the inequality (b) follows from (74). Thus, the inequality (26) holds.

Appendix C: Proof of Theorem 2

According to (26), the matrix $(\mathbf{\Phi}_e)_{\mathcal{T}}$ does not satisfy (2) with probability less than or equal to $4\,(12/\delta_S)^S\, e^{-C_0\lfloor K_e/2\rfloor}$ for any subset $\mathcal{T} \subset \{1, \ldots, N\}$ of cardinality S. Since there are $\binom{N}{S} \leq (Ne/S)^S$ different subsets $\mathcal{T}$ of cardinality S, $\mathbf{\Phi}_e$ does not satisfy the RIP with probability

$$\begin{aligned}
&\Pr\{\mathbf{\Phi}_e \text{ does not satisfy RIP}\} \leq 4\binom{N}{S}(12/\delta_S)^S\, e^{-C_0\lfloor K_e/2\rfloor} \\
&\leq 4\,(Ne/S)^S\,(12/\delta_S)^S\, e^{-C_0\lfloor K_e/2\rfloor} = 4e^{-(C_0\lfloor K_e/2\rfloor - S[\log(Ne/S)+\log(12/\delta_S)])} \\
&\leq 4e^{-\{C_0\lfloor K_e/2\rfloor - C_3[\log(Ne/S)+\log(12/\delta_S)]\lfloor K_e/2\rfloor/\log(N/S)\}} \\
&= 4e^{-\{C_0 - C_3[1+(1+\log(12/\delta_S))/\log(N/S)]\}\lfloor K_e/2\rfloor}.
\end{aligned} \tag{78}$$

Setting $C_4 = C_0 - C_3\left[1 + (1 + \log(12/\delta_S))/\log(N/S)\right]$ and choosing C_3 small enough that guarantees that C_4 is positive, we obtain (27).

APPENDIX D: PROOF OF LEMMA 2

The method of the proof is the same as the one used to prove Lemma 1 and is based on splitting the rows of $\mathbf{\Phi}_e$ into a number of sets with independent entries. Here, the splitting is carried out as shown in (29).

Let $(\mathbf{\Phi}_e)^i_{\mathcal{T}},\ i = 1,\ldots,n_p-1$ be the matrix containing the $(i-1)K+1$-th to the iK-th rows of $(\mathbf{\Phi}_e)_{\mathcal{T}}$. The last $K_e-(n_p-1)K$ rows of $(\mathbf{\Phi}_e)_{\mathcal{T}}$ form the matrix $(\mathbf{\Phi}_e)^{n_p}_{\mathcal{T}}$. Since the matrices $(\mathbf{\Phi}_e)^i_{\mathcal{T}},\ i = 1,\ldots,n_p-1$ consist of independent entries, they satisfy (2) each with probability of at least $1-2\left(12/\delta_S\right)^S e^{-C_0K}$. For the same reason, the matrix $(\mathbf{\Phi}_e)^{n_p}_{\mathcal{T}}$ satisfies (2) with probability greater than or equal to $1-2\left(12/\delta_S\right)^S e^{-C_0K_{n_p}}$. In the event that all the matrices $(\mathbf{\Phi}_e)^i_{\mathcal{T}},\ i = 1,..,n_p$ satisfy (2) simultaneously, for $\boldsymbol{c}\in\mathbb{R}^S$ we have

$$\sum_{i=1}^{n_p}\frac{K_i}{N}(1-\delta_S)\|\boldsymbol{c}\|_{l_2}^2 \le \sum_{i=1}^{n_p}\|(\mathbf{\Phi}_e)^i_{\mathcal{T}}\boldsymbol{c}\|_{l_2}^2 \le \sum_{i=1}^{n_p}\frac{K_i}{N}(1+\delta_S)\|\boldsymbol{c}\|_{l_2}^2$$

$$\Rightarrow \qquad \frac{K_e}{N}(1-\delta_S)\|\boldsymbol{c}\|_{l_2}^2 \le \|(\mathbf{\Phi}_e)_{\mathcal{T}}\boldsymbol{c}\|_{l_2}^2 \le \frac{K_e}{N}(1+\delta_S)\|\boldsymbol{c}\|_{l_2}^2. \tag{79}$$

Therefore, using the union bound and (79), we can conclude that

$$\begin{aligned}\Pr\{(\mathbf{\Phi}_e)_{\mathcal{T}} \text{ does not satisfy (2)}\} &\le \sum_{i=1}^{n_p}\Pr\{(\mathbf{\Phi}_e)^i_{\mathcal{T}} \text{ does not satisfy (2)}\}\\ &\le\ 2(n_p-1)\left(12/\delta_S\right)^S\left(e^{-C_0K}\right)+2\left(12/\delta_S\right)^S\left(e^{-C_0K_{n_p}}\right)\end{aligned} \tag{80}$$

which proves the lemma.

APPENDIX E: PROOF OF THEOREM 3

According to Lemma 2, for any subset $\mathcal{T}\subset\{1,\ldots,N\}$ of cardinality S, the probability that $(\mathbf{\Phi}_e)_{\mathcal{T}}$ does not satisfy (2) is less than or equal to $2(n_p-1)\left(12/\delta_S\right)^S\left(e^{-C_0K}\right)+2\left(12/\delta_S\right)^S\left(e^{-C_0K_{n_p}}\right)$. Using the fact that there are $\binom{N}{S}\le(Ne/S)^S$ different subsets $\mathcal{T}$, the probability that the extended measurement

matrix $\mathbf{\Phi}_e$ does not satisfy the RIP can be computed as

$$\begin{aligned}
\Pr\{\mathbf{\Phi}_e \text{ does not satisfy the RIP}\} &\leq 2(n_p-1)\binom{N}{S}\left(12/\delta_S\right)^S e^{-C_0K} + 2\binom{N}{S}\left(12/\delta_S\right)^S e^{-C_0K_{n_p}} \\
&\leq 2(n_p-1)\left(Ne/S\right)^S\left(12/\delta_S\right)^S e^{-C_0K} + 2\left(Ne/S\right)^S\left(12/\delta_S\right)^S e^{-C_0K_{n_p}} \\
&= 2(n_p-1)e^{-(C_0K - S[\log(Ne/S)+\log(12/\delta_S)])} + 2e^{-\left(C_0K_{n_p} - S[\log(Ne/S)+\log(12/\delta_S)]\right)} \\
&\leq 2(n_p-1)e^{-\left\{C_0K - \frac{C_3K_{n_p}}{K}[\log(Ne/S)+\log(12/\delta_S)]K/\log(N/S)\right\}} \\
&+ 2e^{-\left\{C_0K_{n_p} - C_3K_{n_p}[\log(Ne/S)+\log(12/\delta_S)]K_{n_p}/\log(N/S)\right\}} \\
&= 2(n_p-1)e^{-\left\{C_0 - \frac{C_3K_{n_p}}{K}[1+(1+\log(12/\delta_S))/\log(N/S)]\right\}K} + 2e^{-\{C_0 - C_3[1+(1+\log(12/\delta_S))/\log(N/S)]\}K_{n_p}}.
\end{aligned} \tag{81}$$

Denoting the constant terms as $C_4 = C_0 - C_3\left[1+\left(1+\log\left(12/\delta_S\right)\right)/\log\left(N/S\right)\right]$ and $C_4' = C_0 - \left(C_3K_{n_p}/K\right)\times\left[1+\left(1+\log\left(12/\delta_S\right)\right)/\log\left(N/S\right)\right]$, and choosing C_3 small enough in order to guarantee that C_4 and C_4' are positive, we obtain (31).

REFERENCES

[1] E. J. Candes, and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, pp. 21–30, March 2008.

[2] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, Dec. 2005.

[3] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[4] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, pp. 4036–4048, Sept. 2006.

[5] M. Wakin, J. N. Laska, M.F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K.F. Kelly, and R.G. Baraniuk, "An architecture for compressive imaging," in *Proc. IEEE ICIP*, Atlanta, USA, Oct. 2006, pp. 1273–1276.

[6] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Joint sourcechannel communication for distributed estimation in sensor networks," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3629–3653, Oct. 2007.

[7] Z. Yu, S. Hoyos, and B. M. Sadler, "Mixed-signal parallel compressed sensing and reception for cognitive radio," in *Proc. IEEE ICASSP*, Las Vegas, USA, Apr. 2008, pp. 3861–3864.

[8] G. Taubock, and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *Proc. IEEE ICASSP*, Las Vegas, USA, Apr. 2008, pp. 2885–2888.

[9] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Proc. IEEE CISS*, Princeton, USA, Mar. 2008, pp. 5–10.

[10] Y. C. Eldar, "Compressed sensing of analog signals in shift-invariant spaces," *IEEE Trans. Sig. Processing*, vol. 57, No. 8, pp. 2986–2997, Aug. 2009.

[11] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Sig. Processing*, vol. 57, No. 3, pp. 993–1009, Mar. 2009.

[12] C. de Boor, R. De Vore, and A. Ron, "The structure of finite genetated shift-invariant spaces in $L_2(\mathbb{R}^d)$," *J. Funct. Anal.*, vol. 119, No. 1, pp. 37–78, 1994.

[13] Y. M. Lu and M. N. Do, "A theory for sampling signals from a union of subspaces," *IEEE Trans. Sig. Processing*, vol. 56, No. 6, pp. 2334–2345, Jun. 2008.

[14] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, No. 11, pp. 5302–5316, Nov. 2009.

[15] J. N. Laska, S. Kirolos, M.F. Duarte, T.S. Ragheb, R.G. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *Proc. IEEE ISCAS*, New Orleans, USA, May 2007, pp. 1959–1962.

[16] O. Taheri and S. A. Vorobyov, "Segmented compressed sampling for analog-to-information conversion," in *Proc. IEEE CAMSAP*, Aruba, Dutch Antilles, Dec. 2010, pp. 113–116.

[17] W. Badjwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak, "Toeplitz-structured compressed sensing matrices," in *Proc. IEEE SSP*, Madison, USA, Aug. 2007, pp. 294–298.

[18] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Proc. IEEE Radar Conf.*, Waltham, MA, USA, Apr. 2007.

[19] D. L. Donoho and J. Tanner, "Counting faces of randomly projected polytopes when the projection radically lowers dimension," *Journal of the American Math. Society*, vol. 22, no. 1, pp. 1–53, Jan. 2009.

[20] E. Candes and T. Tao, "Near optimal signal recovery from random projections: universal encoding stategies?," *IEEE Trans. Inf. Theory*, vol. 52, No. 12 , pp. 5406–5425, Dec. 2006.

[21] R. Baraniuk, M. Davenport, R. De Vore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, Jan. 2008.

[22] D. Donoho, "For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution," *Communi. Pure and Applied Math.*, vol. 59, pp. 797–829, Jun. 2006.

[23] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communi. Pure and Applied Math.*, vol. 59, pp. 1207–1223, Aug. 2006.

[24] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.

[25] D. Angelosante, G. B. Giannakis, "RLS-weighted LASSO for adaptive estimation of sparse signals," *Proc. IEEE ICASSP*, Taipei, Taiwan, Apr. 2009, pp. 3245–3248.

[26] C. Craig, "On the Tchebycheff inequality of Bernstein," *Ann. Math. Stat.*, vol. 4, no. 2, pp. 94–102, May 1933.

[27] Y. Zhang, "Theory of compressive sensing via l_1-minimization: A non-RIP analysis and extensions," , 2010.